HIGHWAY RESEARCH RECORD

Number 322

Travel Analysis

17 Reports

HIGHWAY RESEARCH BOARD

NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING

HIGHWAY RESEARCH RECORD

Number 322

Travel Analysis

17 Reports

Subject Areas

55 Traffic Measurements
84 Urban Transportation Systems

HIGHWAY RESEARCH BOARD

DIVISION OF ENGINEERING NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING

WASHINGTON, D.C. 1970

Standard Book Number 309-01821-8

Price: $5.40

Available from

Highway Research Board
National Academy of Sciences
2101 Constitution Avenue
Washington, D.C. 20418

Department of Economics, Finance and Administration

R. C. Blensly, Chairman
Oregon State University, Corvallis

Kenneth E. Cook
Highway Research Board Staff

COMMITTEE ON FREIGHT TRANSPORTATION ECONOMICS
(As of December 31, 1969)

Donald E. Church, Chairman
U. S. Bureau of the Census, Washington, D. C.

Josephine Ayre, Secretary
Bureau of Public Roads, Washington, D. C.

Burton N. Behling
David W. Bluestone
Robert Brown, Jr.
Braxton B. Carr
D. Kenneth Chacey
John E. Clayton
W. A. C. Connelly
Allan C. Flott
Alexander French
Jerome Gilbert
William D. Hart
W. Fletcher Lutz
William L. Maloney
Edward Margolin
William B. Saunders
John E. Stephen
Charles A. Taff
Charles V. Wootan

Department of Urban Transportation Planning

(As of December 31, 1969)

J. Douglas Carroll, Jr., Chairman
Tri-State Transportation Commission, New York

James A. Scott
Highway Research Board Staff

Frank N. Barker
Kurt W. Bauer
Fred J. Benson
Donald S. Berry
Siegfried M. Breuning
E. Wilson Campbell
F. Stuart Chapin, Jr.
John J. Cummings
Leo Cusick
Harmer E. Davis
John G. Duba
Alfred Eisenpreis
John R. Hamburg
Robert L. Hardin, Jr.
Thomas J. Hart
Frederick O. Hayes
Patrick Healy
Bernard F. Hillenbrand
E. H. Holmes
William L. Hooper
Edgar M. Horwood
John T. Howard
Richard Ives
J. E. Johnston
Peter A. Lewis
Mo Chih Li
Burton W. Marsh
J. O. Mattson
Robert E. McCabe
J. B. McMorran
William L. Mertz
Charles L. Miller
Robert B. Mitchell
Clifford F. Rassweiler
Carlton C. Robinson
Joseph L. Schofer
Paul W. Shuldiner
Merlin Smelker
Wilber E. Smith
Francis C. Turner
Alan M. Voorhees
Edward G. Wetzel
C. R. Wilder
F. Houston Wynn
Charles J. Zwick

COMMITTEE ON ORIGIN AND DESTINATION
(As of December 31, 1969)

Alan M. Voorhees, Chairman
Alan M. Voorhees and Associates, Inc., McLean, Virginia

Norman Mueller, Secretary
Bureau of Public Roads, Washington, D. C.

Mark M. Akin
Willard H. Armstrong
Herman Basmaciyan
Austin E. Brant, Jr.
Glenn E. Brokke
Henry W. Bruck
Nathan Cherniack
Donald E. Cleveland
Francis E. Coleman
James H. Cox
Roger L. Creighton
Lewis W. Crump
John A. Dearinger
Thomas B. Deen
Harold Deutschman
John W. Dickey
George A. Ferguson
Michael G. Ferreri
Devere M. Foxworth
W. L. Grecco
Douglas F. Haist
Harold W. Hansen
C. S. Harmon
Philip Hazen
Kevin E. Heanue
Donald M. Hill
Frank E. Horton
Thomas F. Humphrey
Wilbur R. Jefferies
G. H. Johnston
Louis E. Keefer
Norman Kennedy
Dewey Lonsberry
Dana E. Low
Ted Luke
Frank J. Mammano
Brian V. Martin
James J. McDonnell
William L. Mertz
John K. Mladinov
John D. Orzeske
Robbie W. Parker
William S. Pollard, Jr.
Lloyd A. Rivard
James J. Schuster
Arthur Schwartz
Billy J. Sexton
Paul W. Shuldiner
Jacob Silver
Bob L. Smith
Max R. Sproles
Peter R. Stopher
Vergil G. Stover
Edwin N. Thomas
Anthony Tomazinis
Robert E. Whiteside
George V. Wickstrom
David K. Witheford
Martin Wohl
Robert T. Wood
Richard D. Worrall
Robert H. Wortman
J. E. Wright
F. Houston Wynn
Jason C. Yu

COMMITTEE ON LAND USE EVALUATION
(As of December 31, 1969)

John R. Hamburg, Chairman
Creighton-Hamburg, Inc., Bethesda, Maryland

Carl Swerdloff, Secretary
U. S. Department of Transportation, Washington, D. C.

Rodolfo J. Aguilar
Brian Barber
Robert E. Barraclough
David E. Boyce
F. Stuart Chapin, Jr.
Salvatore V. Ferrera
Aaron Fleischer
William L. Garrison
Ronald S. Graybeal
David A. Grossman
Walter G. Hansen
Britton Harris
George C. Hemmens
Donald M. Hill
C. David Loeks
David D. Longmaid
Ira S. Lowry
John K. Mladinov
Kenneth J. Schlager
Morton Schneider
David R. Seidman
Joseph R. Stowers
Robert J. Wheeler
Lowdon Wingo, Jr.
Richard D. Worrall

Department of Traffic and Operations

(As of December 31, 1969)

Harold L. Michael, Chairman
Purdue University, Lafayette, Indiana

E. A. Mueller
Highway Research Board Staff

John E. Baerwald
David M. Baldwin
Donald S. Berry
Murray Blumenthal
Paul C. Box
Siegfried M. Breuning
Donald G. Capelle
William L. Carson
Donald E. Cleveland
Robert E. Conner
Theodore W. Forbes
George J. Gaudaen
Daniel L. Gerlough
Bruce D. Greenshields
Marble J. Hensley
Richard C. Hopkins
George W. Howie
Slade F. Hulbert
Fred W. Hurd
John W. Hutchinson
Charles J. Keese
Alger F. Malo
Garland E. Marple
Ellis L. Mathes
Adolf D. May, Jr.
William A. McConnell
D. Grant Mickle
William J. Miller, Jr.
J. P. Mills, Jr.
Theodore F. Morf
Karl Moskowitz
Joseph C. Oppenlander
Harry Porter, Jr.
Charles W. Prisk
Edmund R. Ricker
Carlton C. Robinson
Matthew C. Sielski
Wilbur S. Smith
Max R. Sproles
Jack Sternbach
K. A. Stonex
Asriel Taragin
Alan M. Voorhees
Edward G. Wetzel
W. W. Wolman

Foreword

Many of the papers contained in this RECORD focus on the need for improved transportation analysis that will assist the engineer and planner in developing transportation plans. The development of appropriate tools that are both reliable and appropriate to the transportation planning process is receiving greater attention by the research community. This is evidenced by the many varied papers in this RECORD that cover the areas of travel analysis and forecasting methodology, choice of mode of travel, and data collection and analysis.

This RECORD will be of primary interest to the engineers and planners specifically charged with the development of comprehensive transportation plans.

Contents

Motor Carrier Data and Freight Modal Split

KENNETH J. DUEKER, Institute of Urban and Regional Research,
University of Iowa, and
ROBERT J. ZUELSDORF, Wilbur Smith and Associates

Collection of commodity flow data from motor carriers is considered in the context of data requirements for freight modal-split models. A detailed examination of techniques for collecting motor carrier commodity flow data casts doubt as to whether collecting data from motor carriers can reasonably be accomplished and whether the data are really what is needed for freight modal-split models. The paper describes an ideal motor carrier commodity flow data set, compares it to data collected by loadometer studies, and describes alternative methods of collecting elements of the ideal data set. The analysis of motor carrier commodity flow data focuses on the State of Wisconsin, in terms of both existing loadometer study methods and regulation of motor carriers.

•ONE OBJECTIVE at the transportation planning and policy level is to use transportation as a means of enhancing economic growth as well as serving existing and anticipated transportation needs. In addition, there is a desire to equitably balance transportation modes and to provide for more efficient utilization of transportation resources. These objectives, however, are usually quite removed from capabilities to anticipate the consequences of alternative policies and plans. Usually these objectives are couched in terms of questions such as, How can transportation resources be effectively developed and efficiently utilized to direct a region's economic growth? How much, for what purposes, and between which points will people travel? Thus, more comprehensive transportation plans are sought to bring together highway, rail, mass transit, port, waterway, and airport development planning. The difficulties of integrating separate mode plans are only now being realized. On the one hand, there is little theory to provide a conceptual framework about commodity transportation; on the other hand, data are lacking about commodity and person flows and the reasons for those flows. Particularly lacking are detailed origin-destination commodity flow data.

With respect to planning for commodity transportation, there appears to be a gap between approaches. The model-builders make what appear to be impossible data demands, whereas data collectors obtain readily collected data rather than data that are useful for model-building.

The primary purpose of this paper is to describe an ideal motor carrier commodity flow data set, to compare it with data now collected by loadometer studies, and to describe alternative methods of collecting elements of the ideal data set. The ideal motor carrier commodity flow data set is one that serves for both modal planning and regulation of motor carriers. A secondary purpose of this paper is to relate the motor carrier commodity flow data to the data requirements for a freight modal-split model and to arrive at tentative conclusions regarding methods for meeting those data requirements.

Analysis of motor carrier commodity flows is used to relate to freight modal-split data requirements because the motor carrier mode would be the most difficult for which to obtain data. The large number and diverse types of motor carriers make it extremely difficult to collect representative flow data. The analysis of motor carrier commodity flows focuses on Wisconsin for specifics in terms of both existing loadometer study methods and regulation of motor carriers.

Paper sponsored by Committee on Freight Transportation Economics and presented at the 49th Annual Meeting.

FREIGHT MODAL SPLIT

This section explores the existing freight modal-split models to determine whether their data requirements can be attained by the various existing commodity flow data collection schemes.

Modal-split models are essential when performing multimodal planning efforts, which are especially important to state Departments of Transportation. The emerging Departments of Transportation (DOT's) should attempt to consider all modes of transport, even those in the experimental stage. The DOT's are attempting to go beyond treating individual modes as self-contained, independent systems and move toward "comprehensive," intermodal transportation planning. Existing methods of estimating the manner in which travel demand will allocate itself between the competing modes (i.e., modal split) have serious limitations.

This section describes two models proposed for freight mode choice estimation and attempts to determine the data requirements for each. The first model described is an abstract mode model developed by Mathematica for the Northeast Corridor Transportation Project (1). The second model was developed for the State of Pennsylvania (2). The Mathematica abstract mode approach predicts demand, by mode, for freight transportation between specified origin-destination pairs. This approach accomplishes the freight generation and distribution tasks as well as performs the modal split. The model predicts the tons of freight to be shipped from node i to node j by mode k in a specified time period T_{kij} by considering eleven variables:

$$T_{kij} = f\left(P_i, P_j, Y_i, Y_j, M_i, M_j, N_{ij}, H_{ij}^b, H_{kij}^r, C_{ij}^b, C_{kij}^r\right)$$

where

T_{kij} = tons shipped by mode k from node i to node j;

P_i, P_j = population of node i, node j;

Y_i, Y_j = gross regional product of node i, node j;

M_i, M_j = industrial character index for node i, node j;

N_{ij} = number of modes available for use between node i and node j;

H_{ij}^b = least time in transit of all available modes between node i and node j;

H_{kij}^r = relative time in transit (with respect to H_{ij}^b) of mode k between node i and node j;

C_{ij}^b = least transportation cost of all available modes between node i and node j; and

C_{kij}^r = relative transportation cost (with respect to C_{ij}^b) of mode k between node i and node j.

Using this approach a mode can be described by its characteristics rather than by its name, allowing for the examination of new modes by merely specifying the travel time and cost of using the mode without referring to its physical form. This mode choice procedure evaluates the effects of transportation time and cost on the modal-split decisions. Cost and time data are useful in evaluating the effects of future transportation investment decisions on the economy and transportation network of a region. The cost of transportation also can be used to predict the distribution of future freight flows.

A disadvantage of the Mathematica freight modal-split model is the lack of consideration given to commodity type. Because of commodity differences in perishability and value density, time and reliability are more important for some commodities than for others. Yet these characteristics are not considered in the model.

Apparently it is necessary to stratify data by commodity type and to develop separate estimating equations for different commodity groupings. Stratification by commodity type, however, requires more point-to-point commodity flow data for model calibration than is presently available.

The model proposed for Penn-DOT performs the modal-split decision using the output of an econometric model. This model provides the amount, in tons of commodity i, that will be shipped from node j to industry j in node X_{igjh}. The problem is to allocate the node-to-node flows of commodities provided by the econometric model to the various modes of transportation. The Penn-DOT model is similar to the Mathematica abstract mode model in mathematical formulation, except that the former has straight estimation of mode choice rather than generation and distribution of freight.

The Penn-DOT modal-split model requires data on the proportion of flow X_{ighj} using each mode. Equally difficult to collect are the data on the flows themselves needed to calibrate the econometric model, which provides the input X_{ighj} to the modal-split model.

For calibration purposes, it is necessary to collect data on commodity types flowing between nodes by the different modes under study. In addition to commodity data on origin to destination by mode, the Penn-DOT model requires stratification of data by shipping and receiving industry type. These requirements impose severe demands on any data collection scheme. An ambiguity in the Penn-DOT model is the distinction between commodity and industry. The model seems to require a one-to-one relationship; i.e., each industry type producing one predominant commodity. In input-output analysis, only dollars of sales between industries are desired. In commodity flow analyses, commodity flow data between these industries are required, which is much more complex. There are various methods of collecting commodity flow data, none of which is totally sufficient. Before the alternative collection methods are discussed, the data to be collected will be outlined.

AN IDEAL MOTOR CARRIER COMMODITY FLOW DATA SET

The ideal data set yields the maximum amount of useful information about commodity flows. It is thought that if the following information could be collected, the major regulatory and planning uses of commodity flow data would be attained.

Identification Requirements

An ideal commodity flow data set will identify the individual motor carrier, the shipper, and the consignee. Each will be identified by name, address, and, in the case of the consignees and shippers, industrial classification. One use of the names and addresses of the three parties will be to supply the means of contacting the parties at a later date if some of the information originally collected is incomplete or inaccurate. This follow-up capacity is necessary to ensure a comprehensive and complete study of all sampled movements. The address therefore must be a mailing one.

Another use of shipper and consignee names and addresses will be to identify the origins and destinations of the goods being transported. As such, the addresses also must indicate geographical groupings by states, cities, or smaller areas.

Transportation Characteristics

Vital to a commodity flow study is information pertaining to the vehicle load itself. An important load characteristic is the identification of the commodities being transported. A classification of commodities must be used that will give the appropriate degree of detail to yield useful information yet is capable of aggregation in various ways for analysis and publication and is precise enough to be collected readily. Various commodity classification systems exist, but quite often the breakdowns are not appropriate for the multiple-purpose objectives desired here.

The Standard Transportation Commodity Code (STCC) is recommended as the best means of classifying commodities (3). The code is revised constantly to keep it up to date. The first five digits of the STCC are identical to the Transportation Commodity Classification used by the Bureau of the Census in its 1963 Census of Transportation; and, as such, the use of the STCC will enable comparison of the collected data with Bureau of the Census transportation data.

The STCC is a rather lengthy document and contains both a numerical and an alphabetical listing of commodities. The great advantage of the code is the hierarchical

form that groups the goods according to varying degrees of detail, the two-digit level being very general with greater detail accruing with each additional digit until at the seven-digit level the specific commodity is described. A study thus can choose the level of detail best suited to its needs and use just that level. Most alternative commodity classification schemes do not offer this ease of detail identification.

The initial commodity classification at the two-digit level breaks all commodities into 35 major industrial groups. The three-digit level lists minor industrial groups, such as jewelry, musical instruments, and toys. The STCC four-digit level differentiates between specific industries. This level divides toys into their various types; games, dolls, children's vehicles, etc. It introduces such items as sporting goods but goes into no further detail as to types of sporting goods. The five-digit level indicates a further breakdown, such as the difference between fishing, hunting, skiing, and football gear. Only the seven-digit level, however, yields a precise description of the many goods transported. For instance, the seven-digit level differentiates between such sporting goods as skis, ski boots, ski bindings, and other skiing equipment.

Because transportation rates and regulations apply directly to the specific commodities (e.g., ski boots) it is believed that any proposed goods movement study should entail commodity detail at the seven-digit STCC level. By utilizing this degree of detail, all commodities sampled will have an identification number. In many past studies, commodity groupings were quite general with the result that one observer would place a commodity in one general grouping whereas another observer would place the same commodity in a different grouping. This problem is minimized by using a very detailed commodity classification scheme such as the STCC.

Another advantage of using great detail in the recording of commodities is that anyone making subsequent use of the data can aggregate the data according to any level of detail desired.

The Standard Transportation Commodity Code is the desired classification because of its superior detail, organization, certainty, and the fact that more and more studies are now using it. A past problem with studies of this type was that two or more studies could not be compared because they used different commodity classification schemes. Wider usage of one system such as the STCC will be a step in eliminating this difficulty.

An ideal commodity flow data set will also collect rate information under which each commodity is being moved. The rates will lend insights into determining the reasons for certain commodity flow decisions; e.g., mode, route, and operating authority used. Rate data are especially useful to the regulatory bodies.

Summary of Ideal Commodity Flow Data

Table 1 gives the set of ideal commodity flow data, some of which have been discussed previously, and identifies the potential users of the data items. The data outlined in this section are the ideal data, and existing data collection techniques may not permit all of these data to be collected. An attempt should be made to collect as much of it as possible, and these ideal data should be the tool for determining which data collection technique is to be implemented. The collection technique producing the data that most closely approximate the ideal data set is the desired technique.

TABLE 1

IDEAL DATA SET

Ideal Data	Planning	Regulation
Identification requirements:		
Consignee name	O	O
Consignee address	X	O
Consignee industrial classification	X	—
Shipper name	O	O
Shipper address	X	O
Shipper industrial classification	X	—
Carrier name	O	X
Carrier address	O	O
Carrier operating authority	X	X
Carrier permit No.	—	X
Vehicle type	X	—
Vehicle identification data	O	O
Vehicle ownership	X	X
Document numbers	O	O
Transportation characteristics:		
Commodity	X	X
Commodity origin-destination	X	X
Vehicle origin-destination	X	X
Dates of flow	X	—
Timing of flow	X	—
Load weight	X	—
Gross weight	X	X
Rates	X	X
Routes taken	X	X

X = data will be used per se.
O = data will be used only to facilitate follow-up to obtain additional data.
— = data not used.

EXISTING LOADOMETER STUDIES

Loadometer studies carried out by state highway departments are the only existing internal studies conducted by states that classify loaded vehicles according to commodities carried. The basic problem with the existing loadometer studies is that the goals of the studies are not the same as the goals and uses of an ideal commodity flow data set. The loadometer studies exist to obtain truck dimension and weight data for highway design purposes, to collect origin-destination data for highway system planning, and to provide an indication of commodity flow data. Throughout the discussion of the existing loadometer studies, problems are presented that indicate why the existing procedures are inadequate to attain the ideal commodity flow data. The State of Wisconsin experience is cited to provide a specific framework.

Description

The Wisconsin loadometer program stops motor carriers at selected points on the highway network, weighs the vehicles, and interviews the drivers. Data are collected from roadside stations and are aggregated to yield information about the entire road network. The following data are obtained in each interview: type of operating authority, vehicle body type, whether or not vehicle is carrying freight, generalized description of commodity, axle spacings, fuel type, origin and destination by state, truck dimensions, axle weights, and operating permit.

Since its inception, the Wisconsin loadometer study has undergone constant expansion and revision from 12 roadside stations in 1942 to the present 39 stations. As seen by the following description, many problems exist with the present system and the ideal data set is not attained.

The existing loadometer study determines the origin and destination of the driver and sometimes the vehicle, neither of which are generally the same as the origin and destination of the commodities carried. Thus, the commodity origin and destination, important elements in the ideal data set, are not obtained.

The commodity carried is registered at the five-digit STCC level, not the more detailed seven-digit level. The source of the commodity identification is the driver, who in many instances does not know exactly what commodity is carried or who may be carrying a mixed load in which only one commodity or sometimes no commodity is identified.

The vehicle is registered as either full or empty, with no load factor given. Vehicle type, carrier operating authority, gross rate, and date of flow are collected in the loadometer study; the remainder of the ideal data set as given in Table 1 is not collected. The existing study thus does not nearly comply with the ideal data set.

Many other operational and practical problems exist with the present system.

With 39 fixed stations, many roads and road segments are never measured by the loadometer study. Statistical sampling of road segments is not employed, making it difficult to make inferences about total truck traffic on state roads.

A problem occurs at those loadometer study locations using Motor Vehicle Division weighing stations, which are used for overload enforcement. Each driver knows where each fixed Motor Vehicle Department weighing station is and will avoid the station if he has an overload or will go by it only when it is closed. Knowledge of whether the scales are open or closed can be obtained at the nearest truck stops, from other truckers, and even from the trucking companies themselves. Because many vehicles successfully avoid the scales (the exact proportion is unknown), the loadometer sample at the scale suffers. Consequently, the loadometer study, which does not prosecute overloads, is tied to the Division of Motor Vehicles, which does prosecute; and the sample is thus biased by the drivers that avoid the stations.

In 1968, the earliest loadometer station opened on June 10 and the last station closed on September 6. Because the loadometer stations are open only during the summer months, the data collected pertain only to that period. The volume of freight, however, and the types of commodities carried vary during the course of the year. The Wisconsin Department of Transportation recognizes these differing seasonal characteristics and

volumes and consequently does not attempt to project a fall, winter, or spring commodity or vehicle flow figure from the summer data. Data are only available therefore for summer commodity movements.

The times and dates of operation of each loadometer station are also very limited. The vast majority of the stations collect loadometer data only 1 day per year, and the specific dates vary from year to year. The collection is further limited in that the stations are open only on weekdays, thereby collecting no weekend or holiday data. The majority of loadometer stations, furthermore, are open only 8 hours, from 6:00 a.m. to 2:00 p.m. Thirteen of the stations (one-third of the total) are also open from 2:00 p.m. to 10:00 p.m. Only four stations are open at night, 10:00 p.m. to 6:00 a.m., and these are all on the Interstate Highway System.

When a loadometer station is open, the crew attempts to stop, weigh, and interview 100 percent of the trucks going past the station. At many stations, however, during periods of high traffic volume, the physical facilities do not allow a great backup of vehicles. The average stoppage time for a truck is 2 minutes. When it is not possible to stop all trucks passing by, the vehicles are sampled according to a sampling model. The model surveys all five-axle vehicles and samples the smaller vehicles that pass by, thereby eliminating undue congestion. The use of the vehicle sampling plan adds an error factor to the data collected, but is justified by the roadside collection technique and the physical capacities of the stations.

Summary of Loadometer Study Analysis

As seen, the existing data collection technique does not nearly meet the requirements of the ideal data set. Most of the desired data are not collected, and some that are collected are not accurate or complete. The sparcity of stations, the lack of adequate sampling of times of operation, the inflexibility of operations, and the sole reliance on driver interviews all cause problems that require improved techniques to obtain the ideal data set.

ROADSIDE DATA COLLECTION TECHNIQUES

A roadside field survey requires a means of accumulating commodity vehicle flow data directly from vehicles on the roads. The existing loadometer study is an example of this technique. The information can be collected at the roadside by questioning the driver, by examining the documents carried on the vehicle, or by actually inspecting the load within the vehicle. The other roadside method is to identify specific vehicles passing given roadside stations, either with manual tabulation or an automatic device, and then to contact the carriers to obtain the desired information pertaining to the observed vehicle. Each of these techniques is described in the following paragraphs, and the advantages and problems of each are identified. Several problems common to all of the roadside techniques are described first, followed by the problems unique to each technique.

All roadside survey techniques require the establishment of observation or interview locations on road segments in the study area. It is impossible, however, to locate a roadside survey station on all highways and roads in the state. Further complicating the location problem is the ability of vehicles to use any combination of road links rather than staying on one specific identified road. Consequently, to obtain complete data on all highway commodity flows, observations must occur at a statistically significant number of randomly selected locations.

With probability as a technique, the statistician can strike an economic balance between (a) the cost of great precision and detail, e.g., cost of numerous stations, and (b) the losses that arise from insufficient information or precision, e.g., too few stations or poorly located stations. Given a confidence level and a desired degree of accuracy, the sampling theory is used to indicate the portion of the total universe needed to estimate the true commodity flow for a given area and commodity strata. In addition, it may be appropriate to stratify by road types, such as arterial, collector, and local roads. The length of each of these road types in total miles can be ascertained, and a probability sample for each can be developed that will locate the appropriate number of observation stations on each type.

Ideally, all observation stations will record the flows of all motor carriers passing when the station is in operation. Past experience, however, indicates that a study that stops motor carriers for even the simplest questions (e.g., the existing loadometer study) detains vehicles for an average of 2 minutes. If extensive commodity flow data are collected by driver interview, load inspection, or document investigation, the time is likely to be increased. When motor vehicle traffic is sparse, it is perhaps possible to stop and obtain data from all vehicles passing the station, and this should be done. Some stations, however, will be on heavily traveled routes; and during periods of peak flow it is physically impossible to stop all the vehicles. Consequently, vehicles will have to be sampled during these periods.

Load Inspection Method

One method of obtaining commodity information at a roadside survey is to actually open the trailer and examine the commodities being carried. This method, however, is an unlikely prospect. First, nearly all companies maintain a policy that all trailers be sealed and that the seal remain intact in transit. Second, it will take a prohibitive amount of time to enter the trailer and examine each load. Third, through visual inspection it is quite often impossible to identify commodities in a load, especially when the vehicle is carrying many different types of freight. Actually examining the load therefore is not a feasible method.

Vehicle Observation and Follow-Up Method—This technique is to record the license number or trailer identification number of vehicles passing the roadside station and to obtain the carrier's name and address by means of carrier license files. The desired data for each vehicle observed are then obtained from the carrier home office or terminal. No information other than vehicle identification data is obtained at roadside. All motor vehicles display a license and an identification number, and these can be observed readily by a camera or an observer situated along the road. Using this method facilitates a large sample size because of the ease in which the original data can be collected. Great problems exist in obtaining data, however, after the movement has occurred. (See the "Carrier-Contact Collection Techniques" section for a discussion of these problems.)

Driver Interview Method

This roadside survey method involves stopping the vehicles and interviewing the driver, as is now done in the loadometer studies. Many problems exist with this type of survey, some of which stem from the limited knowledge and sometimes uncooperativeness of the drivers themselves. Most drivers never see the commodities they are carrying, so that the only way the driver will be able to identify commodities carried is to look at the freight bills or to reiterate what the dispatcher or dock worker told him he is carrying. Quite often the driver's knowledge of what his vehicle is carrying is general and inaccurate. Unless he looks at the freight bills, a driver will never know what commodities are in the vehicle if the load is mixed, as is the case in one of every six loads in Wisconsin. Therefore, commodity data for these mixed loads will have to come from the freight bills carried with the load. Another problem is that the driver has no idea of the origin or destination of the commodities but rather only knows the origin of his particular trip and where his trip will end. This origin and destination may not be the actual origin or destination of either the vehicle or the commodity. Hence, any method of collecting commodity-flow data by merely questioning the driver of the vehicle must be examined critically, for the data received will be inexact and usually incomplete.

Document Examination Method

A roadside survey technique that does not seem to have been tried in any existing study is to obtain commodity information directly from documents carried in the motor vehicles. The obvious prerequisite for such a technique is the existence of documents containing the necessary information. The documents carried in motor carriers

operating on highways depend on the type of carrier—common, contract, or private. According to the Wisconsin Administrative Code:

> (1) Freight bills for each shipment handled shall be made and kept by the carrier showing the name and address of the carrier, consignor, and consignee; the origin and destination; the date of receipt by the carrier; the description by number of packages and commodity name; and the weight, rate, and charge. (2) On traffic moving under joint rates, freight bills shall also show the point of interchange, the name of the connecting carrier, and the division of revenues between the joint carriers. (Public Service Commission, PSC 16, 02 Bills, freight.)

It appears from the regulations that intrastate common motor carriers need not carry the actual freight bills in the vehicle. Through extensive interviewing of common carriers, however, it is believed that most, if not all, common carriers do carry the appropriate freight bills in the vehicle carrying the freight.

Each interstate common carrier must also issue a document that will contain the following information, as required by the Interstate Commerce Commission (4): date of shipment; names of consignor and consignee; points of origin and destination; number of and description of packages; exact description of articles; and weight, volume, or measurement on which charges apply.

Again, all common carriers in interstate commerce seem to carry this document (either a freight bill or a bill of lading) in the vehicle that contains the freight.

All but the exempt contract carriers, both interstate and intrastate, must maintain bills of lading (which contain some of the ideal data for a commodity flow study) and must have these bills in the vehicle when carrying the commodities. The exempted contract carriers are those that carry farm and forest products directly from the point of growth or production. Although no bill of lading or freight bill need be carried on the exempted carriers, it appears that these carriers do carry a bill of sale that usually names the commodity and lists the weight, the origin and destination (shipper and purchaser), and the price paid. Hence, there appears to be little problem with contract carriers not carrying adequate documents from which commodity flow information can be obtained.

Private motor carriers by law must carry some document or have other means of proving that the commodities carried in the vehicle are actually owned by the carrier. They need not carry freight bills or bills of lading, although most private carriers do carry a bill of lading of sorts. The reason for this is that the company itself wants to know exactly what commodities are in each vehicle. Consequently, most companies (certainly the larger firms with more than one vehicle) do not allow a vehicle to leave a terminal or other place of origin without an appropriate bill of lading. Often the private vehicles are actually large delivery vehicles that deliver the commodities to the company's outlets; e.g., chain food or department stores. These carriers carry detailed documents enumerating the commodities within the vehicle.

Very few vehicles, therefore, move on the highway without bills of lading, freight bills, or receipts of sale from which commodity flow data can be obtained. When a vehicle does not have a document from which to derive the commodity information, the possible procedure is to obtain the carrier's name and address; record the license number and fleet number of the truck; and follow up by a mail questionnaire, phone call, or interview to the carrier home office or terminal to obtain the data about the particular vehicle.

Several problems, however, do exist with this method of examining documents in the vehicle. It is easy to deal with the documents of a load containing one or two commodities and therefore one or two freight bills, which can be manually copied or photostated. Many vehicles, however, carry mixed loads and many freight bills. In the case of common carriers, the number of freight bills on each vehicle can vary from one to well over 100. Consequently, either a large staff must be maintained at each station, the bills must be photostated, or the bills must only be sampled.

A problem in obtaining data directly from freight bills carried on the vehicle is the accuracy and nature of the data on the bills. Table 2 gives the data that

can be obtained directly from the freight bill. From Table 2 it is apparent that almost all the ideal data can be obtained by using interviewer observation to supplement the freight bill data. Problems with the freight bill data that require clarification are of the following types: vehicle origin and destination are by company terminal number rather than address, no mailing address for consignee, and a typing error (which might state "ch supplies" instead of "sch supplies" for identifying the commodity). Arthur D. Little (5) found that the description of commodities on freight bills is also a major source of error; for example:

Commodity Shown on Freight Bill	Proper Commodity Description
1 crtn bolts	1 crtn bolts, iron or steel
10 bxs gloves	10 bxs gloves, leather
1 dr oil	1 dr peanut oil
1 box Autolite	1 bx spark plugs

Because of the sources of error on freight bills, follow-up may be required to yield all the ideal data set.

Summary of Roadside Data Collection Techniques

This section has identified the four basic types of roadside data collection: driver interview, inspection of the load, the recording of vehicle identification with a follow-up procedure, and document inspection. The last technique yields by far the most data, but in itself is insufficient to yield the complete ideal set. Hence, a combination of the methods will be required, the most probable being the driver interview and document inspection.

CARRIER-CONTACT COLLECTION TECHNIQUES

Another source of data is the motor carrier home office or terminal. Many problems exist, however, with obtaining information from carrier document files. First, no regulation states where the files are to be kept, thereby leaving the location to the discretion of the carriers. Small carriers will maintain the files at the home office or terminal. Large carriers may also do this, or the documents may be kept at the terminal if origin or destination. If the latter is done by a significant number of carriers, it will be extremely difficult to obtain the data from each terminal; a state study would have to contact every terminal in the country. It may be possible, however, to obtain cooperation from the carrier in easing this problem.

A second problem is that the method of filing the documents is up to the individual carrier. The filing systems vary between sophisticated cross-reference files to drawers filled with old documents. Generally, each carrier maintains its files alphabetically according to the consignee's or shipper's name. Such a filing system does not lend itself to yielding bills according to vehicle or shipment date because to supply such bills the entire alphabetical files would have to be paged through, as is the case with contacting shippers. This is possible for large carriers that make use of computers to record their

TABLE 2

DATA OBTAINED FROM FREIGHT BILLS IN ROADSIDE SURVEYS

Ideal Data	Collection Results
Identification requirements:	
Consignee name	A
Consignee address	B
Consignee industrial classification	B
Shipper name	A
Shipper address	A
Shipper industrial classification	B
Carrier name	A
Carrier address	C
Carrier operating authority	A
Vehicle type	D
Vehicle identification data	A
Vehicle ownership	C
Document numbers	A
Transportation characteristics:	
Commodity	B
Commodity origin-destination	A
Vehicle origin-destination	B
Dates of flow	D
Timing of flow	D
Load weight	D
Load factor	C
Gross weight	D
Rates	A
Routes taken	D

A = data completed on freight bill.
B = data on freight bill required clarification.
C = data not obtainable.
D = data obtained at roadside to complement document data.

data, but the small carriers would have to do the work manually. For this reason, carriers would have to be instructed in advance concerning the sampling procedure so that the bills could be intercepted prior to filing. The third problem with using carrier files is that some carriers, the exact proportion unknown, will not have files. This proportion, however, is not expected to be large, and obtaining the data prior to filing will eliminate the problem.

A very major problem with any technique that entails direct contact with the carriers themselves is the task of identifying the carriers that operate on Wisconsin roads. At present, no list exists of all motor carrier firms or motor vehicles that operate within the state. Such a list would have to be compiled prior to any study that would involve contacting the carriers.

SHIPPER-CONTACT COLLECTION TECHNIQUES

It is possible to contact shippers by means of a questionnaire, telephone call, personal interview, or by requesting that they send in a certain sampling of their shipping documents. Each of these contact techniques requires that data be collected from shipper documents; and, as such, the document mailing technique may be the least expensive and most accurate. Certain problems, however, cause great difficulty in requesting that sampled shippers send in copies of freight and shipping bills. Even though most commercial shippers maintain shipping document files from which the desired documents can be obtained, the files are usually kept alphabetically according to consignee's name. Because of such filing methods the desired documents cannot be gathered for all commodities shipped on a specific date or in a specific vehicle without first paging through the entire year's shipping documents. Such a document search would be most difficult for most shippers. Second, most small shippers keep very limited files and some, e.g., farmers and individual small firms, keep no useful files at all.

The greatest difficulty with any method of contacting shippers is the magnitude of the effort that would be involved in contacting all shippers of goods into, out of, within, or through a study region such as the State of Wisconsin. Industrial and commercial firms, the largest shippers, can be obtained from several different lists. These shippers, however, will have to be recorded for the entire United States because much of the freight carried on Wisconsin roads originates with out-of-state shippers. All shippers thus must be contacted; or, more likely, a sample must be taken of all shippers throughout the United States. Even if this were possible, many shippers do not know if the goods went through a given state or not. A commodity flow study for a single state or region must find a better source of data or be part of a national study.

If a universe of shippers can be established and sampled and if a means of intercepting a sample of shipment documents is developed, the shipper-contact collection technique should be extremely viable at the national level. An input-output study collects interindustry dollar flows. An extension of this would be to collect commodity and mode information. More importantly, data in this form are more suitable for direct input to freight modal-choice models.

SUMMARY OF DATA COLLECTION METHODS

Many methods of collecting commodity flow information exist that might be useful in a commodity flow study. Table 3 gives the types of data that can be obtained from each method. All the ideal data cannot be collected; however, several techniques are better than others. In selecting a final technique, the data each method is capable of collecting are very important. The wealth of data, however, will have to be balanced against the costs in time and money in obtaining the information. All the following methods should also play a role in selecting the data collection technique.

Any method must be statistically sound and the sampling method must be spread over the entire year, with all dates having an equal probability of being sampled, so as to isolate all possible adverse effects of sampling.

The least possible burden must be placed on the carriers or shippers. All data must be obtained in the first contact with the carrier or shipper, with follow-up procedures to be implemented only when absolutely necessary. Detailed step-by-step

TABLE 3

SUMMARY OF SOURCES AND DATA COLLECTED

Ideal Data	Data Sources: Contact Carriers — Documents Mailed In	Contact Carriers — Terminal Visits	Roadside Techniques — Inspect Documents	Roadside Techniques — Driver Interview	Roadside Techniques — Load Inspection	Roadside Techniques — Vehicle ID and Follow-Up
Follow-up data:						
Carrier name	X	X	X	X	X	X
Consignee mailing address	X	X	O	O	O	X
Consignee and shipper names	X	X	X	O	O	X
Consignee and shipper addresses	X	X	X	O	O	X
Document number	X	X	X	O	O	X
Vehicle number	X	X	X	X	X	X
Load information:						
Commodity description	X	X	X	O	O	X
Commodity origin-destination	X	X	X	O	O	X
Load weight	O	X	O	X	X	X
Load factor	O	X	O	X	X	O
Load type (mixed, straight)	O	X	X	X	X	X
Number of commodities	O	X	X	O	O	O
Commodity weight	X	X	X	O	O	X
Rate	X	X	X	O	O	X
Vehicle data:						
Route	O	X	O	X	O	O
Dates arrival, departure	O	X	O	X	O	X
Times in transit	O	X	O	X	O	X
Vehicle origin-destination	X	X	X	X	O	X
Carrier operating authority	X	X	X	X	X	X
Vehicle ownership	O	X	O	O	O	X
Interlining	X	X	X	X	O	X
Industrial classification of shipper-consignee	O	O	O	O	O	O

X = data can be collected. O = data cannot be collected.

instructions must be supplied to all sources of data, so that all necessary personnel can cope with every circumstance. A process of quality control to ensure accuracy and completeness of data must be implemented. Each carrier must also be guaranteed anonymity, with absolutely no adverse effects accruing on data sources.

CONCLUSIONS

In looking at the most difficult mode of collecting commodity flow data—surveying motor carriers—it can be concluded that collecting data from carriers is extremely complex and costly. The problems involved, as discussed in the analysis, are immense. Great difficulty is encountered in approximating the ideal motor carrier commodity flow data set because of the variety of carriers and commodities on the highways. These difficulties raise considerable doubt as to whether surveying carriers can provide a single data set useful for both planning and regulation.

For purposes of a freight modal split, it appears that contacting shippers should be investigated in greater detail. In a broader framework, such as a national or large regional study, the determination of a universe of shippers from which to sample might be less difficult.

REFERENCES

1. Mathematica. Studies on the Demand for Freight Transportation. Prepared for the Northeast Corridor Transportation Project, U. S. Department of Transportation, Washington, D. C., August 1967.
2. Methodology Framework for Comprehensive Transportation Planning. Prepared for the Governor's Committee for Transportation, Commonwealth of Pennsylvania, Technical Advisory Committee Project 6(b), by Transportation Research Institute, Carnegie-Mellon University and Pennsylvania State University (undated).

3. Standard Transportation Commodity Code. Association of American Railroads, Washington, D. C., January 1, 1967.
4. Motor Carrier Information Bulletin No. 1. Interstate Commerce Commission, October 1966, p. 6.
5. Little, Arthur D. Pilot Study Report. Transportation Division Data Bank, Public Utilities Commission, Case 66460, August 1965, p. B-8.

Passenger Terminal Impedances

JEFFREY M. BRUGGEMAN and RICHARD D. WORRALL, Peat, Marwick, Mitchell and Co.

Conventional transportation network analysis requires the estimation of the times necessary to complete three portions of a typical intercity trip. Although the times associated with the line-haul and access portions have been studied extensively, the time required for transfer between access and line-haul modes has not been handled adequately. These times, or impedances, occur at any intercity passenger terminal for air, rail, and bus modes. This paper identifies each of the major components of the passenger terminal system and develops the respective impedance methodology for each as well as a technique for combining them into a single value representative of the total impedance level for a particular terminal. Data were collected at several intercity terminals in the Washington, D. C., area; and the impedance levels were determined for each terminal. The methodology and many of the component values are directly transferable to other terminals. In addition, the methodology may be used as an aid in evaluating alternative functional arrangements of the various terminal facilities.

•A CONVENTIONAL network simulation model distributes transportation flows across the network as an inverse function of the difficulty, or "impedance," of travel along each link. For a regional network, at least three major categories of impedance may be defined:

1. Link impedance is the measure of the average "line-haul" travel times, costs, or distances associated with travel along the major links of the regional network.
2. Access impedance is the measure of the average "access" times, costs, or distances involved in traveling from a typical origin or destination point to the nearest access/egress point on the regional network.
3. Terminal impedance is the measure of the average transfer (or "terminal") times, costs, or distances involved in transferring from the access system to the line-haul system.

For a typical air journey from Washington, D. C., to Boston, the link impedance might be represented by the average air travel time between Washington National and Logan International Airports; the access impedance, by the respective travel times from the original starting point in Washington to Washington National Airport and from the airport in Boston to the final destination; and the terminal impedance, by the respective times spent within the airports in Washington and Boston.

To date, the attention of model-builders has focused on estimates of link impedances and, more recently, access impedances. Little or no attention has been devoted to estimates of terminal impedances. This paper describes a first attempt to develop a system of terminal impedance measures for incorporation into the existing Northeast Corridor network simulation model(s).

In this study, attention is directed primarily toward time measurements, covering the period between the arrival of the passenger in the line-haul terminal to his departure by either the line-haul or the access/egress mode. The estimates are based on the output of a series of simple queueing models embedded in a matrix of estimated walking times. Data for the study were developed from "as built" plans of existing Corridor

Paper sponsored by Committee on Origin and Destination and presented at the 49th Annual Meeting.

terminals and were supplemented by limited field studies of walking speeds and terminal process times developed in the Washington, D. C., area.

In addition to providing terminal time estimates as inputs to a simulation model system, the materials developed in this study also serve as a framework within which to evaluate the operational efficiency of alternative functional arrangements within the terminal and as a convenient empirical device for identifying current focal points of delay.

IMPEDANCE METHODOLOGY AND DEFINITIONS

Three separate measures of terminal time were initially developed for possible inclusion in the network model system. The first, termed "average elapsed time," represents the total time spent by the average traveler within the terminal system, measured from his instant of entry to his time of exit. This measure was rejected for a variety of reasons. First, such a measure would include time spent in various nonterminal activities such as eating a meal or shopping in gift shops. Second, such a definition does not lend itself readily to mathematical modeling. Finally, data collection would require passenger contacts, a technique to be avoided in crowded terminals.

A second, more meaningful definition, termed "average terminal time," is used extensively in the following analysis. The average terminal time is made of three separate elements: the "processing time," the "engaging/disengaging time," and the "movement time." Processing time represents the average time taken by a passenger to perform specific, travel-related terminal functions, such as ticket purchase or baggage checking. Engaging/disengaging times are the times at which a passenger may be assumed to have transferred between two of the terminal subsystems discussed in a following paragraph. The difference between an engaging/disengaging time and the time when a specific event occurs (e.g., the time of the traveler's arrival at the boarding area before scheduled departure time) is used as an element of the average terminal time. Finally, the movement time is simply the time required to move from one process or engaging/disengaging point to another.

The third measure, "minimum essential processing time," represents the least amount of time that a typical passenger must spend within the terminal system. This value differs from the average terminal time in that the engaging/disengaging times are replaced by the equivalent processing times associated with the engaging/disengaging activity.

For departing passengers, separate average terminal time and minimum essential processing time estimates are developed for each of the terminals studied. For arriving passengers, however, the two definitions produce essentially the same results; thus only one estimate is required and is computed for the average passenger. For passengers transferring from one line-haul vehicle to another, a minimum essential processing time may be estimated for assumed typical interchanges.

The operation of a passenger terminal may be generalized as shown in Figure 1. Three types of flows are indicated: line-haul departures, line-haul arrivals, and line-haul transfers. For analytical convenience, the terminal has been broken down into three subsystems; an access mode transfer system, a terminal processing system, and a line-haul transfer system. As the three types of flows have somewhat different needs in each of three subsystems, a matrix of nine subsystems is formed. For all practical purposes, the termination of the line-haul mode system will be the same for both line-haul arrivals and line-haul transfers, and the initiation of the line-haul mode system will be the same for line-haul departures and line-haul transfers. A total of seven different subsystems thus may be examined.

Each of these subsystems may be broken down into many separate components. The example shown in Figure 2 illustrates the operation of the terminal processing system when used for line-haul departures. This chart is applicable for use in determining both average transfer times and minimum essential processing times. Passengers are received from the termination of the access mode system either with or without baggage. Each group may be further broken down into these who are preticketed and those who must acquire their tickets at the terminal. A further breakdown of those groups with

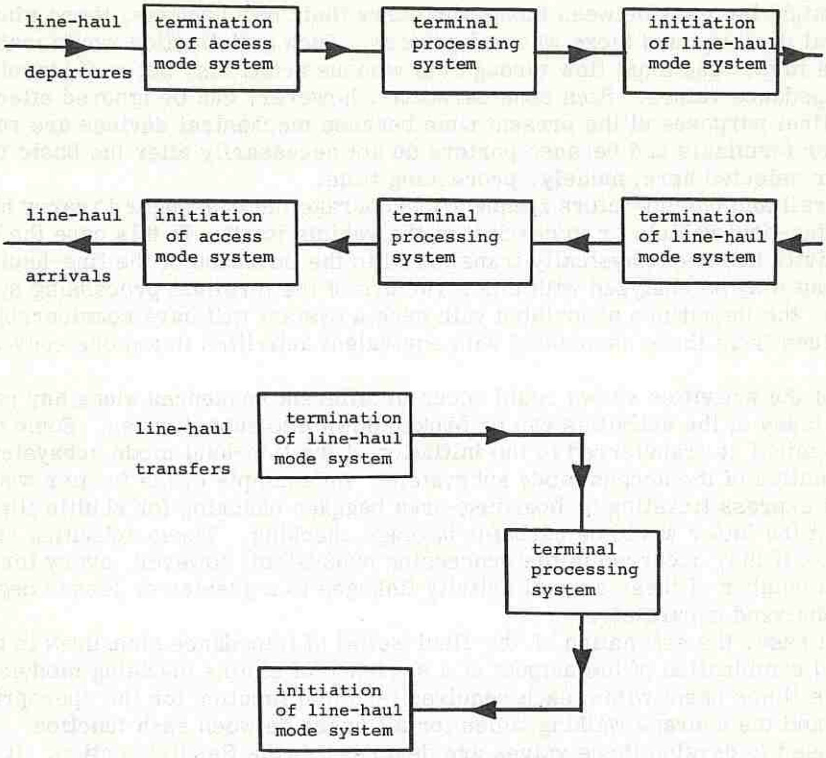


Figure 1. Terminal processing system.

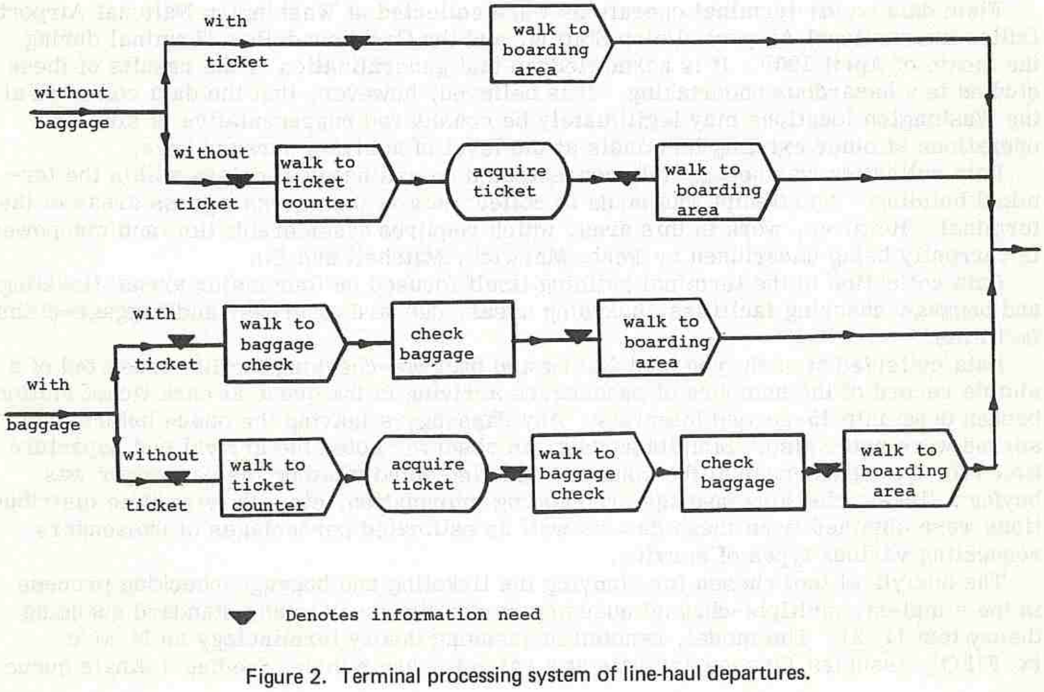


Figure 2. Terminal processing system of line-haul departures.

baggage might be made between those who carry their own baggage, those who use mechanical devices, and those who use porters. Such a distinction would make little difference in the functional flow through the various activities, but could involve different impedance values. Such considerations, however, can be ignored effectively for analytical purposes at the present time because mechanical devices are rarely used at corridor terminals and because porters do not necessarily alter the basic impedance parameter selected here; namely, processing time.

Also, rail and bus operators frequently encourage the passenger to carry his baggage onto the line-haul vehicle or to check it at the vehicle itself. In this case the baggage-check activity has been physically transferred to the initiation of the line-haul mode system, but may be analyzed with this structure of the terminal processing system. Obviously, the impedance associated with such a system will have considerably different values from those associated with equivalent activities in a more conventional system.

Most of the activities shown could occur in different sequences along any path. In addition, many of the activities can be broken down into subactivities. Some of the activities might be transferred to the initiation of the line-haul mode subsystem or to the termination of the access mode subsystem. An example of the former would be the use of express ticketing or boarding-area baggage checking for shuttle flights. An example of the latter would be curbside baggage checking. These activities can be analyzed as if they occurred in the processing subsystem; however, every terminal contains a number of these special activity linkages to a greater or lesser degree and must be analyzed separately.

In each case, the estimation of the final set(s) of impedance measures is based on a weighted combination of the outputs of a sequence of simple queueing models (representing the times spent within each required terminal function for the appropriate set of users) and the average walking times for all users between each function. The actual methods used to develop these values are described in the Results section. It should be noted that the estimates made do not consider explicitly the actual paths through the terminal, the perceived versus actual delays, or the congestion effects.

DATA COLLECTION AND ANALYTICAL PROCEDURES

Field data on air terminal operations were collected at Washington National Airport, Dulles International Airport, Union Station, and the Greyhound Bus Terminal during the month of April 1969. It is acknowledged that generalization of the results of these studies is a hazardous undertaking. It is believed, however, that the data collected at the Washington locations may legitimately be considered representative of similar operations at other existing terminals at the level of analysis pursued here.

Data collection was necessarily concentrated on estimates of delays within the terminal building. No attempt was made to collect data in the access/egress areas of the terminal. However, work in this area, which requires considerable time and manpower, is currently being undertaken by Peat, Marwick, Mitchell and Co.

Data collection in the terminal building itself focused on four major areas: ticketing and baggage-checking facilities, boarding areas, deboarding areas, and baggage-claim facilities.

Data collected at each type of ticketing and baggage-checking facility consisted of a simple record of the numbers of passengers arriving in the queue at each ticket station, broken down into 15-second intervals. Any passengers leaving the queue before being served were noted also. Simultaneously, an observer noted the arrival and departure times for each passenger at the counter, and also noted whether the passenger was buying a ticket, checking baggage, requesting information, etc. Service time distributions were obtained from these data as well as estimated percentages of passengers requesting various types of service.

The analytical tool chosen for studying the ticketing and baggage-checking process is the simplest, multiple-channel queueing model discussed in any standard queueing theory text (1, 2). The model, denoted in queueing theory terminology as M/M/c: (∞/FIFO), assumes Poisson arrivals at a rate of λ per minute, feeding a single queue

from which all stations or service channels are then fed. Although this model clearly does not apply exactly to the situation found in most airport ticket areas, the "jockeying" of passengers from one queue to another results in an effective queue discipline not unlike that assumed by the model. The assumption of exponentially distributed service times with mean value μ is similarly a reasonable approximation to the situation for most types of service.

Arrival times at the boarding gate were recorded by simply counting in successive 15-second intervals the number of passengers arriving over a period of 30 to 40 minutes before the scheduled departure of the aircraft. Data were collected only for those flights that had no posted departure delays. Additional time spent in the terminal because of in-flight delays is not to be considered within the purview of this study as terminal impedance. In the boarding area, similar observations were made of the processing times at the check-in desk and the rate at which passengers boarded the aircraft. Information concerning ticket purchase at the gate, gate baggage checking, and standby passenger handling was recorded along with the check-in times in a manner analogous to that used to record the supplementary data collected at the ticket counters in the main terminal.

Observations of arriving passenger flows included the rate at which the passengers passed out of the aircraft door. Separate studies were made of flow rates for different types of unloading devices. The type of device (see the Results section) proved to have a significant effect on the deplaning rates. It should be noted also that observations at Dulles International were made at the exit of the mobile lounge rather than at the aircraft door because the scheduled time for flights to and from that airport includes the necessary travel times via the mobile lounge from the terminal to the loading apron. Likewise, boarding-area arrival and boarding-gate studies at Dulles were all made at the mobile lounge entrance.

The fourth and final set of observations was made in the baggage-claim area. In this case a single observer recorded the arrival rates of baggage arriving from a given flight at the claim device. Simultaneously, one or more other observers recorded the rates at which this baggage was picked up by the passengers. These observations were coordinated with the passenger deplaning data to obtain the previously discussed estimates of the line-haul disengaging time for passengers with baggage.

The results of the analyses performed on these data were combined with estimated walking distances between terminal functions scaled from plans of airports. In accordance with the definitions given earlier, straight-line minimum paths were traced. The probability of significant passenger deviation from these paths, particularly because of a lack of adequate directional information, is acknowledged but is not treated explicitly in the analysis because the intent simply is to generate a simple, gross estimate of average terminal time. The walking distance to the boarding area is an average measure for all gates with no allowance for weighting by gate utilization. Several paths were traced for different airlines at Washington National Airport, where substantial asymmetry exists between various carriers.

These distances were converted to average walking times using an assumed average walking speed of 4 feet per second based on data derived from previous studies of walking speeds (3, 4). Further stratification of walking speeds, based for example on age/sex breakdown, was thought to be an unwarranted refinement at the current level of analysis. Congestion effects at specific points such as doorways and stairs likewise were ignored. The effect of these assumptions is not critical in the final impedance measure.

Similar data were collected at Union Station and at the Greyhound Bus Terminal and from plans of these facilities. Data for both regular train and Metroliner service were collected at Union Station because separate facilities are provided for these services and because passenger behavior was found to be somewhat different. No such breakdown was required at the bus station.

RESULTS

This section focuses on the results obtained from the field studies at Washington National Airport and uses these findings to develop a set of simple important measures

for this terminal. Equivalent findings for Dulles International Airport, Union Station, and the Greyhound Bus Terminal are summarized at the end of the section.

Ticketing and Baggage Check

Many different arrangements of ticket and baggage-checking facilities are in use by different airlines. One major distinction is the degree to which different service requests—e.g., ticketing and information—are separated and assigned to different counter stations. Different service combinations produce radically different service time distributions at a given counter station.

In most cases a primary, advanced-reservation ticket counter is provided, together with specialized express baggage, will-call, and information stations. Inevitably, however, some passengers approach the counter area at the wrong location, requesting a type of service not provided at that position. Airline policy, generally, is to serve these passengers whenever possible rather than to redirect them to a more appropriate counter. The relative inefficiencies introduced by this approach are, of course, more than balanced by its customer relations value.

The service time distributions observed at the primary ticket counters for four different airlines at National Airport are shown in Figure 3. Figure 3(a) shows the results of a highly specialized operation. Although express baggage and special information counters are provided to attract many of these brief service transactions, a substantial number of relatively short service times are still observed at the primary counter.

Figure 3(b) shows the distribution of service times for another airline that operates extensive express baggage and special service stations, again in addition to the primary ticketing station. This distribution differs substantially, however, from that shown in Figure 3(a). The airline in question operates a relatively simple set of routes from National Airport, involving fewer connections and less selection between alternatives than does that shown by Figure 3(a). As a result, the average service times are much shorter and the service time distribution is more compact.

Figure 3(c) shows an operation where provision is made for baggage checking at facilities away from the ticket counter. Again, a fairly complex routing structure leads to some quite lengthy service times.

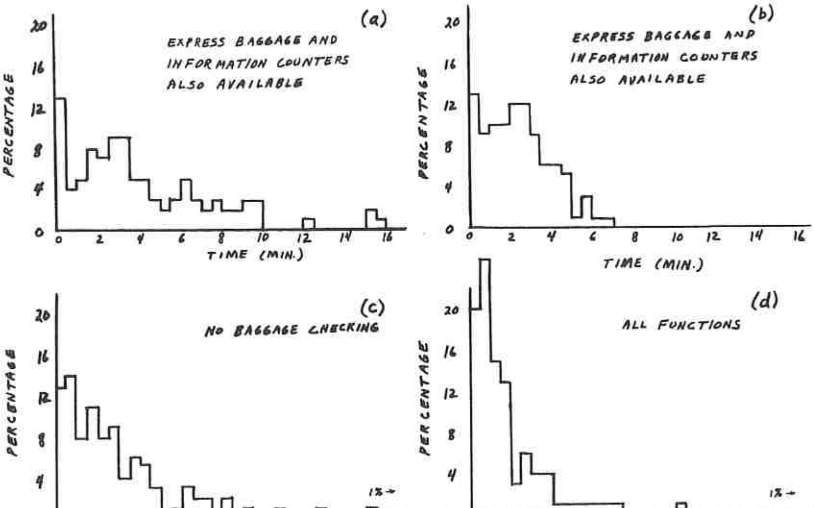


Figure 3. Washington National Airport ticket counter service times.

Finally, Figure 3(d) shows a situation where virtually all activities—ticket, baggage checking, information—are concentrated at a group of undifferentiated counters. The effect of brief information and baggage checking services is quite pronounced, whereas the airline's simple routing structure from National Airport serves also to eliminate the majority of complex and lengthy bookings.

The operation of an express baggage counter, shown in Figure 4(a) for one typical airline is essentially similar to that of the information counter 4(b) with the exception that extremely short times are somewhat less frequent. Again, there is a limited number of relatively long service times, reflecting transactions involving passengers with unusually large quantities of baggage.

The distribution of service times associated with a separate information counter is shown in Figure 4(b). Over half of the service times are of less than 30 seconds duration; i.e., they would fall within the first time intervals shown in Figure 3. A limited number of somewhat longer service times was also observed.

The service time distributions for the ticket counters and the baggage-checking counters at Dulles International Airport are shown in Figures 4(c) and 4(d) respectively. Very few extremely long service times were observed. The relatively large proportion of very short service times indicates that a considerable number of simple information requests were handled at the primary ticketing counters, despite the presence of the special information counter. The similarity between the two distributions illustrated in Figures 4(c) and 4(d) is partly explained by the occurrence of a considerable amount of "cross-servicing"; i.e., tickets were purchased at the baggage counter and bags were checked at the ticket counter. In fact, except during brief peak surges, the available differentiated counters were used effectively for undifferentiated service.

The major purpose of specialized service counters is to provide rapid service to those passengers with simple requests, to reduce overall passenger delays, and to reduce the load on those counter agents who must handle the more complex and lengthy transactions. An efficient alternative, or more properly an efficient supplement, to a system of specialized counters is the use of a "floating server" or agent in the lobby area ahead of the counter. These persons contact passengers as they approach a queue

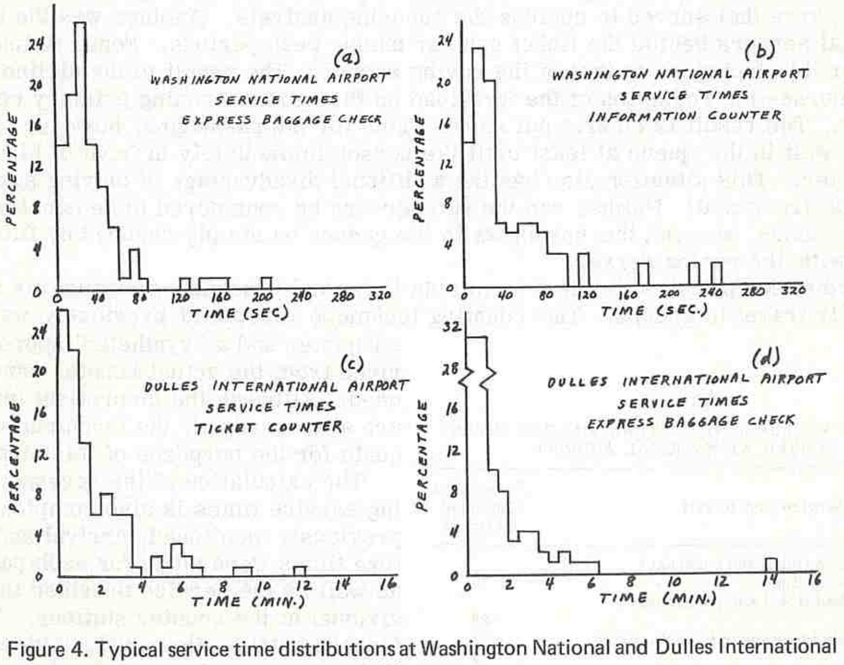


Figure 4. Typical service time distributions at Washington National and Dulles International Airports.

and provide service when possible or direct the passenger to the proper line when more specialized attention is required. This activity is called by various names by different airlines—front counter service, random servers, roving servers, etc. The latter term will be used here.

One of the most effective roving-server operations at Washington National Airport is carried on by Eastern Airlines. In this case, one or more roving servers are employed during all busy periods. They operate either by contacting the passenger and eliciting information on the service he requires or by simply responding to passenger queries, which are usually simple information requests and often are not connected with Eastern's operations at all. Under normal circumstances, the agent will answer information requests or confirm reservations while the passenger is waiting in line. In unusual cases, such as a last-minute passenger needing a ticket changed, he either will take the passenger to a vacant counter station and handle the request himself or else will lead him to the front of an existing queue and request that the agent deal with his request next. (This latter practice occurs very infrequently, and does not seriously affect the assumptions underlying the queueing models discussed here.)

The roving server provides an extremely valuable service to certain travelers by considerably reducing their waiting times and, of course, generates considerable "good will" for the airline. His main function is to reduce the level of input to the queue, to prevent the occurrence of major unnecessary delays, and to reduce congestion and confusion in front of the counter area.

Eastern Airline's roving servers at Washington National Airport concentrate on passengers queueing at the primary ticket counter. A limited study showed that roughly 32 percent of all passengers arriving at these counters were both contacted and served in this fashion. In addition, about 1 percent of the passengers in the express baggage line were similarly served. These figures represent a conservative evaluation of the efficiency of the system in that they do not include passengers served before they joined a specific queue.

The times for roving service are naturally very short. Roughly two-thirds of all service times were of less than 30 seconds duration, although an occasional lengthy contact was made. These latter were typically rush services of the type mentioned before. A simple breakdown of the service provided by the agent is given in Table 1.

The presence of the roving server was just one of several complications in the ticketing area that served to confuse the queueing analysis. Another was the use of additional servers behind the ticket counter during peak periods. Some airlines seemed to prefer this technique to that of the roving server. The result to the airline is similar of course—the reduction of the workload on the agents manning primary counter stations. The result is clearly not so desirable for the passenger, however, because he must wait in the queue at least until the person immediately in front of him reaches the counter. This situation also has the additional disadvantage of defying simple analytical treatment: Neither can the two servers be considered to be handling independent queues, nor can the net inputs to the queues be simply reduced by filtering as occurs with the roving server.

A further complication lay in determining the actual passengers because air travelers frequently travel in groups. The counting technique discussed previously was not adequate, and a "synthetic" approach derived from the actual counter services was used. Although the theoretical inadequacies are acknowledged, the technique was adequate for the purposes of this study.

TABLE 1

SERVICES PROVIDED BY EASTERN AIRLINES ROVING SERVER AT NATIONAL AIRPORT

Service and Result	Passengers Contacted (percent)
Agent checked ticket and passenger remained in line	39
Agent checked ticket and passenger left line	33
Agent directed passenger to another line	22
Other	6

The calculation of the necessary queueing service times is also complex. As previously mentioned, arrival and departure times were noted for each passenger as well as the service provided that individual at the counter stations. The difference between these values gives an estimate of the service. A simple average of

TABLE 2

QUEUEING TIMES AT WASHINGTON NATIONAL AIRPORT

Airline	Activity	Waiting Time (min)	Service Time (min)	Total Time (min)
A	Ticketing	2.94	4.05	6.99
A	Express baggage	5.59	0.54	6.13
B	Ticketing	4.57	2.74	7.31
B	Express baggage	1.78	0.57	2.35
C	Ticketing	1.56	2.96	4.52
C	Baggage	4.18	0.80	4.98

these latter times, however, does not provide the necessary input to the queueing model because a few seconds of agent bookkeeping time typically occur between the end of one service and the beginning of the next. This extra time must be included in the model to give a true service-start to service-start block time.

Once the average input rates and the average service rates are available, the average time spent waiting in the queue may be determined directly from standard, published charts for the appropriate queueing model. See, for example, Molina (5) and Lee (1). To this waiting time must then be added the average service time obtained from the raw data to yield an estimate of the average total delay to a typical passenger. The results obtained from such an analysis for three typical airlines at National Airport are given in Table 2.

Boarding Area

A basic input to the impedance calculations is an estimate of the time before scheduled departure that the passenger commits himself to the line-haul system by entering the boarding area. Distributions of passenger arrivals were obtained from the field data for a representative sample of departing flights, and the time for the average (50th percentile) passenger was computed. These average times and the distributions themselves were found to vary considerably both among flights and among airlines.

At least five different factors influence the shape of these distributions.

1. The time of day—passengers tend to arrive closer to the scheduled departure time for flights leaving early in the morning or at the close of the business day than for midday flights.
2. Destination of the flight—all other things being equal, a flight to a location associated with pleasure or vacation travel, such as Miami, attracts passengers earlier than a heavily business-oriented flight; e.g., Washington to New York.
3. Frequency of service—closely allied to the destination factor is the passenger's perception of the penalty associated with missing a flight. Service for some flights is much more frequent, and the consequent penalty for missing a flight is much less severe than for others.
4. Number of connecting passengers—the impact of connections on the distribution of arrivals at the boarding area depends, of course, on the meshing of the connecting flight schedules. The most noticeable effect of passengers arriving from a connecting flight is a sudden surge in the arrival rate. This can have significant impact on the average value, particularly from one day to another, if the on-time performance of the connecting flights varies.
5. Availability of an attractive holding area—again other things being equal, this is probably one of the most significant variables for seasoned travelers; if the passenger knows that the boarding area is particularly uncomfortable to wait in, he is much more likely to spend any excess time at some other place in the terminal than if an attractive, comfortable waiting area is available.

Figure 5(a) shows the arrival pattern for three different flights on the same airline at Washington National Airport. As might be expected, the arrival patterns suggest that the New York flight, with its associated high proportion of business travelers, has the "tightest" distribution (i.e., the largest proportion of late arrivals) whereas the passengers tend to be more "strung out" for the Memphis-Dallas flight. Although it cannot be proved directly from the data, one might suspect that the early arrivals for the Chicago flight are pleasure or vacation travelers headed for connecting flights at O'Hare Airport, whereas the last-minute stragglers, following the New York pattern, are businessmen.

Also shown in Figure 5(a) is a curve developed by Paullin (6) for San Francisco International Airport. Other flights, not shown here, were also plotted against Paullin's curve. Some, particularly long-distance flights, showed much better agreement than those shown by Figure 5. Others, however, were steeper than the New York flight and thus deviated further from Paullin's curve.

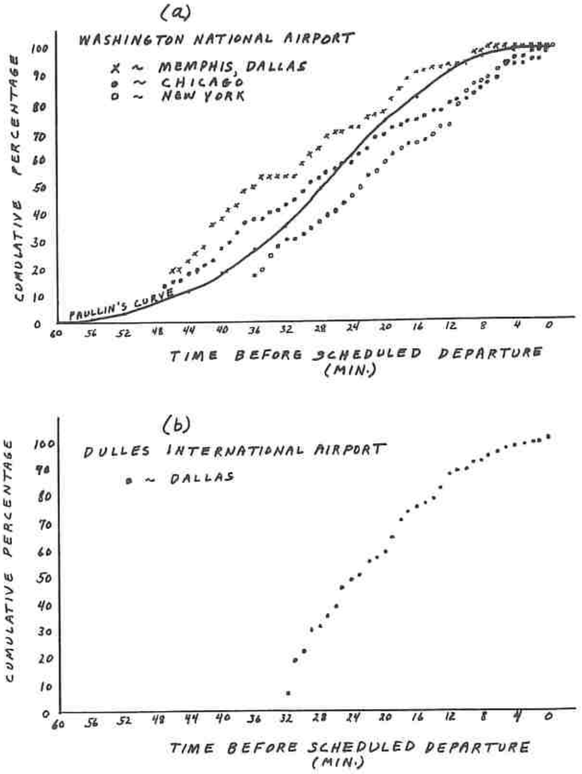


Figure 5. Airline passenger arrival patterns.

Data on passenger arrival distributions at the Dulles mobile lounge boarding area were taken for a variety of flights. A typical flight is shown in Figure 5(b). The most remarkable thing about this distribution, as compared to those observed at National Airport, is the absence of the traditional S-shaped arrival patterns. The curve for Dulles is, in fact, concave downward throughout its length. This phenomenon may be explained by the reactions of passengers to boarding announcements. Virtually no passengers were found to queue at the lounge until the first boarding announcement was made. Then a very rapid surge of passengers was observed, followed by a gradual tapering off, much as had been observed for conventional boarding areas. A behavior pattern of this type is believed to apply to any terminal where a departure hold room has not been specifically assigned to a flight.

Data were not collected on boarding rates into the lounge because an extensive study of this and other features of mobile lounge operations had recently been completed by the Bureau of National Capital Airports (BNCA). Also, this event occurs after the time a passenger is assumed to be committed to the line-haul system, and thus the event is excluded from the impedance calculation.

The service time distributions at the check-in point or entry point to the boarding area were developed at National Airport. As might be expected, this is a very compact distribution because the range of activities to be carried out by the boarding-area agent is very limited. The distribution times for check-in services varied slightly between airlines. One important source of variation was the use of seat assignment at the check-in desk. As might be expected, flights with seat assignment resulted in longer average check-in times. A few airlines, however, had sufficient flights of each type so that meaningful comparisons could be drawn. Where entries in both columns appear in Table 3, the seat-selection data are usually the result of observations of a single flight, whereas the others are averages computed for several flights.

TABLE 3

BOARDING AREA PROCESSING TIMES AT WASHINGTON NATIONAL AIRPORT

Airline	Seat Selection (min)	No Seat Selection (min)
A	—	0.37
B	0.67	0.52
C	0.44	0.36
D	0.25	0.23

Boarding-area processing times were observed also at Dulles Airport. The processing time over several carriers and flights was found to be 0.24 minute. This did not include

processing for flights for which check-in and seat selection had been done at the ticket counter. For such flights, the boarding process time was trivial; the passengers simply flashed their boarding passes at the gate attendant as they boarded the mobile lounge.

Observations of comparative boarding rates were also made for each of the major types of boarding device currently in use. Because the act of boarding occurs after the passenger has committed himself to the line-haul system, its evaluation is not required for the terminal impedance analysis; these data will not be analyzed in detail here. The data were collected mainly for use in an evaluation of the efficiency of alternative boarding schemes proposed for future research.

TABLE 4

DEBOARDING RATES

Device	Rate (pass./min)
Jetway	32.0
Jetway with stairs	31.7
Weighted average	31.9
On-aircraft stairs	22.1
Mobile stairs	28.9
Weighted average	25.3

Deboarding and Baggage Claim

Data on arriving line-haul passengers were collected at both the deboarding gate and in the baggage-claim area. Passenger deboarding rates were recorded for a range of aircraft types and passenger unloading devices. Average values are given in Table 4 and should be compared with airline industry standards of 25 passengers per minute. The superiority of the jetway is clearly seen from these data; the passengers deboard much faster through this device than if they are required to descend a flight of stairs. The use of a jetway with adjustable stairs within it, such as that used by Eastern at National Airport, does not noticeably impede the passengers. Deboarding rates for this device are virtually the same as for the second-level jetways used by American, for example.

A second interesting point is the significantly slower deboarding rates for aircraft using attached stairways, such as the DC-9 and the 737, as compared to a set of mobile stairs, either self-propelled or mounted on a truck. The ladder-like stairs of these aircraft do not seem easily traveled by passengers and result in a significantly longer and more cumbersome deboarding process.

Only relatively smooth deboarding processes were used to generate these values. Frequently, a considerable delay was generated by one particular individual such as an elderly person or a woman carrying an infant. These delays were much more frequent when stairs were used rather than jetways; thus, the potential difference in the efficiency of the two devices is underestimated rather than overestimated by the data.

Data collected in the baggage-claim area at National Airport proved to be, again, somewhat awkward when using simple manual techniques. Recording the arrival of baggage onto the claim device and coordinating this with aircraft arrival is straightforward. Observing the collection of individual pieces of baggage as they are claimed is less easy, however, in part because several of the claim devices are shared by more than one airline, each of which may have one or more scheduled flights arriving in a short time interval. Thus, baggage from more than one flight often was observed on a single claim device at one time, making proper determination of the claiming activity extremely awkward.

Average claim times for those airline serving Washington National Airport are as follows:

Airline	Baggage Claim Time (min)
A	8.93
B	9.75
C	9.53

These values represent the time after the aircraft door opened when the average (50th percentile) bag was claimed, assuming that the passenger claimed it as soon as

possible; bags left in the claim area 30 minutes after flight arrival were not included in the average.

Deboarding data for Dulles used the results of the BNCA study referred to earlier. The average deboarding time found in this study was 1.90 minutes, and this value was employed subsequently in the analysis.

Data in the baggage-claim areas were again collected for a variety of flights and carriers. The average time interval between the arrival of the mobile lounge at the terminal and collection of the average (50th percentile) piece of baggage was 5.12 minutes. It must be noted, however, that all data collection took place during moderately busy rather than during peak times when the existing baggage delivery process becomes seriously overloaded.

Eastern Airlines Shuttle Service

The Eastern Airlines shuttle service between Washington and New York differs markedly from conventional service. Departing passengers using the shuttle may bypass the main terminal area completely and proceed directly to the boarding area where they pick up a boarding pass from an automated ticket machine. Tickets are purchased on board the plane after it is airborne. Passengers may also purchase shuttle tickets at the special shuttle desk on the lower concourse. The numbers of passengers choosing this option, however, was small during the course of the field study.

Assuming that a passenger simply walks to the boarding area and picks up a boarding pass, the only values required as input to the average transfer time are the walking distance and the time before scheduled departure that the passenger arrives in the boarding area. This latter value, however, is somewhat difficult to determine for all but the early morning flights because the guaranteed-seat service with hourly departures and almost inevitable second sections during busy periods results in a relatively continuous stream of arrivals over the greater part of the day. On the basis of limited observations at Washington National Airport, a value of 22 minutes was estimated as the mean "before departure" arrival time.

For the minimum essential processing time estimates, a nominal value of 30 seconds may be allocated to allow for acquiring the simple boarding pass. Likewise, a value of 10 seconds was assigned for baggage handling at the boarding gate—no checking is required, bags are merely deposited on a device near the boarding gate.

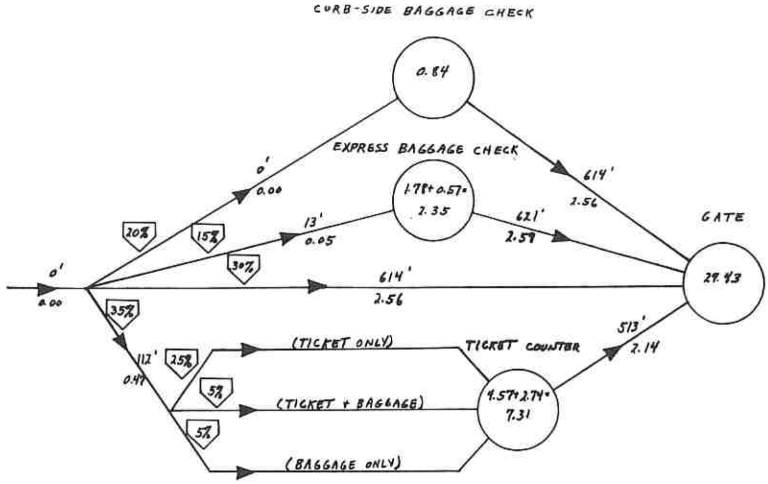
Shuttle passengers arriving at National follow a pattern much similar to that used by passengers on regular system flights. The only significant difference is the use of a separate baggage-claim area and a slightly different average walking distance. Limited data on baggage claiming yielded a value of 11.06 minutes as an average for the 30 percent of passengers estimated to carry baggage on shuttle flights.

Impedance Calculations

The framework for the impedance calculations for a typical airline at National Airport is shown schematically in Figure 6 illustrating the calculation of average terminal time for departures. Other paths would be drawn for other airlines that would reflect the different arrangements of ticketing and baggage-checking options used by each. Schematic drawings likewise could be used to summarize the minimum essential processing time for departure and the average terminal time for arrivals.

Very few passengers purchased tickets or checked baggage at the gate, although the option was available from all of the airlines. The "penalties" associated with each of these activities thus were determined from a very small number of observations. A value of 1.25 minutes was estimated for the ticket purchase penalty, 0.10 minute for baggage checking, and 1.30 minutes for both. Although only crudely estimated, the accuracy of these measures is considered adequate because of the small number of passengers involved and the relatively short times.

The individual processing elements used in the calculations have been discussed earlier. To arrive at a weighted average, however, the "path split" or the percentage of passengers utilizing each of the separate process sequences must be estimated.



$$\bar{T} = 0.00 + 0.20(0.00+0.84+2.56) + 0.15(0.05+2.35+2.59) + 0.30(2.56) + 0.35(0.47+7.31+2.14) + 29.43$$
$$= 35.10 \text{ MINUTES}$$

Figure 6. Calculation of average terminal time departures.

Because of data collection limitations, it was not possible to determine exactly what every passenger was doing as he passed through the terminal. Reliable percentage figures thus could not be determined from field observations alone. An approximate method was therefore used to estimate these splits based on airline passenger flow data. The dangers inherent in this approach are acknowledged, and it is proposed that the validation of these percentages by more rigorous data collection techniques should be a major goal for future research. The final estimated terminal impedance values for Washington National Airport are given in Table 5.

The equivalent calculations of impedances for Dulles International Airport also are given in Table 5. The path splits, or percentages of passengers using each option, are the same as those developed for National Airport. (This assumption is admittedly somewhat tenuous.) No data were collected on carriers using the apron loading facilities at Dulles. It is thought, however, that reliable impedances could be estimated for these carriers by combining the appropriate process times and the walking distances from the terminal plans.

TABLE 5

SUMMARY OF IMPEDANCE VALUES

Process	Average Terminal Time (min)	Min. Essential Processing Time (min)
National Airport:		
Line-haul departures	25.55-35.10	3.15
Line-haul arrivals	7.12	7.12
Shuttle departures	24.59	2.95
Shuttle arrivals	5.39	5.39
Dulles Airport:		
Line-haul departures	24.54	4.17
Line-haul arrivals	1.01	4.17

Summary of Results at Union Station and Greyhound Bus Terminal

Figure 7(a) shows the service time distribution for coach ticketing for all trains at the main ticket counter. This distribution follows a quite typical pattern: A few extremely short times (0 to 10 seconds) appear; the majority lie in the 10- to 80-second range; and a few remaining times are scattered over the 80- to 300-second range. These service times include a number of information requests, which predictably fall primarily in the 0- to 20-second range.

Figure 7(b) shows service times at the Metroliner ticket counter in the main terminal. This distribution is quite widely

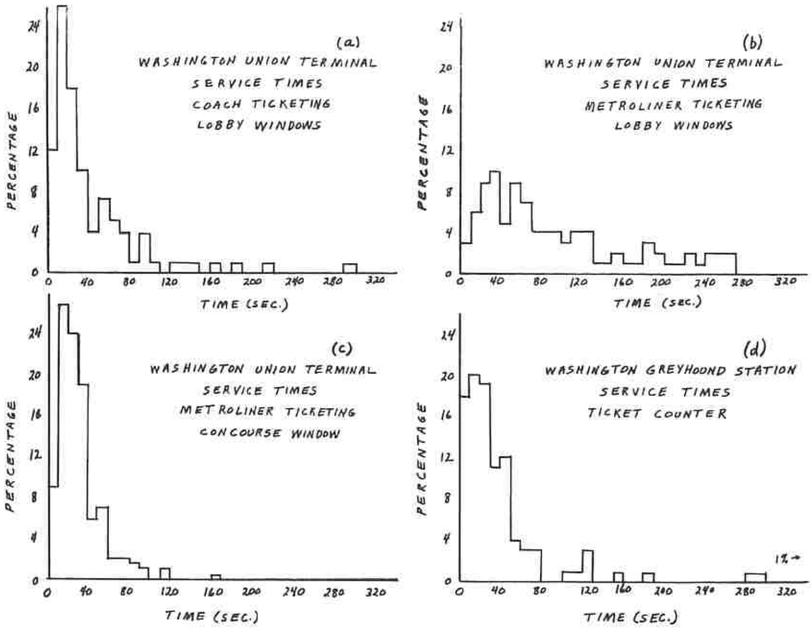


Figure 7. Typical train and bus service time distributions.

spread with a number of service times somewhat in excess of 3 minutes. This result should be interpreted with considerable caution. The ticketing system in use at the time of the study was an interim one, designed to operate only until the planned computerized ticketing and reservation system is installed. Once such a system is in operation, the average processing time should drop radically. It should also be noted that although certain windows are marked expressly for Metroliner service, tickets on other trains may be purchased there. Several complicated bookings for non-Metroliner trains were recorded at these windows, increasing considerably the dispersion of the data and resulting in an effective overestimate of the true Metroliner service time.

Finally, the service times at the special concourse Metroliner window are plotted in Figure 7(c). These times follow a pattern similar to that for the regular coach sales, but with a shorter range. This is readily explained by the fact that any ticket sales at this point represent a definitely available seat, and thus no complex telephoning ahead is required.

The service time distribution at the Greyhound ticket windows in shown in Figure 7(d). The plot shows, as might be expected, a decline in service time frequencies up to 80 seconds followed by a long "tail," representing the occasional lengthy, complex, ticket purchase. Slightly more than 50 percent of the services recorded in Figure 7(d) were for ticket purchase. The majority of the remainder were requests for some type of information, and transactions such as making change also were observed. Most of these nonticketing services were relatively brief, less than 30 seconds duration in most cases.

The results of applying the simple M/M/C: (∞/FIFO) queueing model to rail and bus ticketing are given in Table 6.

TABLE 6

BUS AND TRAIN TERMINAL TIMES

Location	Waiting Time	Service Time	Total Time
Coach tickets, all trains	4.43	0.83	5.25
Metroliner, terminal	1.19	1.84	3.03
Metroliner, concourse	1.71	0.50	2.21
Greyhound station	2.13	0.70	2.83

Note: All times in minutes.

Departing Metroliner passengers tended to arrive at the boarding gate a considerable time before the scheduled departure of the train. A typical arrival pattern for a Metroliner departure is shown in Figure 8(a). The average arrival time for all Metroliner departures studied was 18.61 minutes before departure. This figure was consistent for all trains.

Departing passengers for regular train service, however, showed a markedly different pattern, which is also shown in Figure 8(a). Passengers tended to arrive later for these trains, with an average time of 13.37 minutes, over 5 minutes less than for the Metroliner. This result is somewhat surprising and may be explained by the relative novelty of the Metroliner service. The value will probably decrease as passengers become more familiar with the operation of the service.

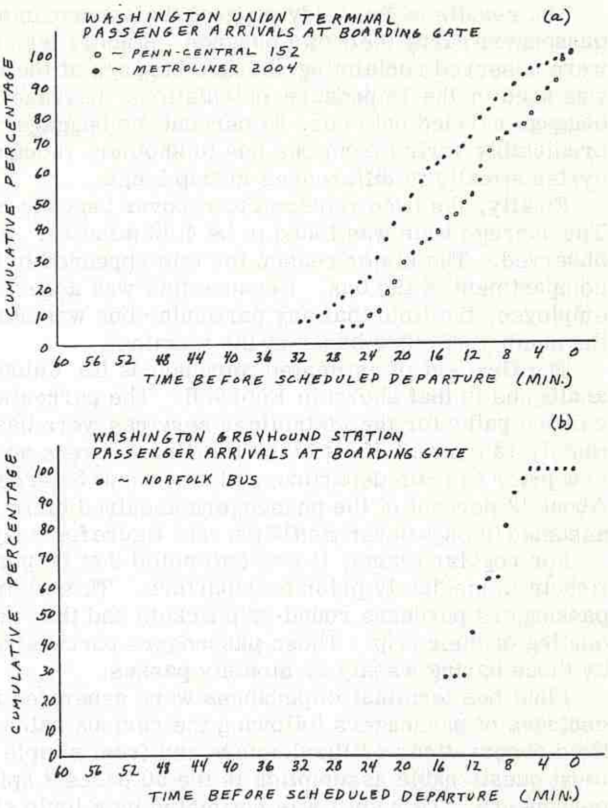


Figure 8. Bus and train passenger arrival patterns.

Departing bus passengers also arrive at the departure gate some time before the bus is scheduled to depart. Their arrival pattern prior to a typical bus departure is shown in Figure 8(b). The first entry on this graph represents the beginning of the data collection period; the earliest arrivals are often found waiting 30 to 40 minutes before the scheduled departure time. The average observed arrival time prior to all departures was 13.50 minutes; variations from the average were related to the size of the busload and the time at which the boarding announcement was made over the public address system.

Arriving train passengers reach the end of the platform at an average time of 3.17 minutes after the train has stopped for the Metroliner and 4.11 minutes for regular trains. Passengers deboarding below the main concourse, of course, must ascend to the concourse level before leaving the terminal. The additional delay thus incurred has been estimated by increasing the necessary walking distance to the concourse gates. No attempt has been made to weight the average walking distance based on tract utilization; a simple arithmetic average has been used. The error involved here is believed to be slight.

No significant difference was observed between the arrival rates of passengers with and without baggage. This indicates that passengers with baggage do not take longer to reach the gate than passengers not so encumbered.

Separate deboarding distributions were recorded for bus passengers with carry-on baggage and those with no baggage. It was found that passengers with baggage required a somewhat longer time to deboard. The median deboarding time after bus arrival for a passenger with baggage was 2.11 minutes, compared to only 1.56 minutes for a passenger without baggage. Whether this difference in fact represents a delay inherent in having to manipulate baggage or simply reflects an element of courtesy in allowing unhindered passengers to leave first is debatable.

The results of the study suggest that approximately 35 percent of deboarding bus passengers carry their own baggage. Slightly less than half of the remaining passengers were observed reclaiming checked baggage at the bus side. The following breakdown was used in the impedance calculations: baggage checked at bus side, 30 percent; baggage carried onto bus, 35 percent; no baggage, 30 percent. These percentages predictably varied from one bus to another, though in no way that could be related systematically to differences in trip length.

Finally, the time required to recover baggage checked at the bus side was recorded. The average time was found to be 4.80 minutes. Again, a substantial variation was observed. The major reason for this appeared to be the delay in opening the baggage compartment of the bus. Because this was done for all buses, by a single terminal employee, the time that any particular bus was serviced was obviously a function of the number of other buses in the terminal.

The final set of estimated impedances for Union Station were calculated in a manner analogous to that shown in Figure 6. The percentages of passengers following the various paths for the Metroliner services were based on the observation that approximately 12 percent of all Metroliner passengers acquired tickets at the concourse window prior to train departure and that about 3 percent picked up boarding passes there. About 12 percent of the passengers acquired their tickets at the lobby window, and an assumed (though unverified) 3 percent figure for boarding-pass pick-up seems reasonable.

For regular trains, it was estimated that 50 percent of the passengers acquire their tickets immediately prior to departure. This figure seems reasonable, since many passengers purchase round-trip tickets and thus need to use a ticket counter only on one leg of their trip. Those passengers purchasing one-way tickets would be offset by those having weekly or monthly passes.

Final bus terminal impedances were generated in a similar fashion. The percentages of passengers following the various paths were determined jointly from the field observations outlined before and from simple visual estimates. Probably the most questionable assumption is the 50 percent split between ticketed and nonticketed passengers. This split was estimated by a logic similar to that given for rail passengers. These values for the Greyhound Bus Terminal given in Table 7 include within-terminal times only. They do not include any time spent in access/egress activities.

EXTENSIONS AND APPLICATIONS

The impedance analysis could be extended in at least three important areas. First, in-depth investigation of passengers reaction to perceived versus actual terminal times could be performed. These investigations would show whether the various impedance components are simply linearly additive or if some elements should be weighted to reflect significant perceived impedances.

Second, the simple queueing model used in this study could be examined and possibly refined. At the same time, other analytical refinements could be investigated such as the development of explicit functions to account for increased delays caused by congestion within the movement areas of the terminal.

Finally, the impedance analysis could be extended into the access/egress areas of the terminal complex. Peat, Marwick, Mitchell and Co. is currently pursuing this topic in its attempt to estimate the delays associated with parking lot operations at major airports. Delays involving other access/egress modes can be estimated in a fairly straightforward manner.

TABLE 7

TERMINAL TIMES

Process	Average Terminal Time (min)	Minimum Essential Processing Time (min)	
		With Ticket	Without Ticket[a]
Union Station:			
System departures	18.31	3.17	9.43
System arrivals	5.83	—	—
Metroliner departures	20.90	2.81	5.07
Metroliner arrivals	4.52	—	—
Greyhound Bus Terminal:			
Intercity departures	15.37	0.64	3.62
Intercity arrivals	3.25	—	—

[a]Rail passengers without tickets are assumed to purchase them in the terminal. If tickets are purchased on the train, the impedance becomes the value given for passengers with tickets.

In addition to its use in estimating impedances associated with the nodes in a transportation network analysis, the methodology discussed in this paper may be applied to the evaluation of alternative functional arrangements within the terminal system. Such alternatives as the provision of purpose-specific ticketing and baggage-checking facilities may be evaluated by the simple queueing models. Other more substantial changes in passenger-processing philosophy, such as the widespread use of computer-printed tickets acquired at remote locations, could be analyzed within this framework; however, individual process time estimates would have to be developed for any novel elements. The evaluation of alternative arrangements thus would become a simple "pen and paper" simulation. This approach is more appropriate for many applications than the development of large-scale, general-purpose computer simulation models.

SUMMARY

A simple methodology for estimating the impedances associated with passenger terminals has been developed and applied to terminals in the Washington metropolitan area. The techniques may be extended readily to other terminals. This would require a minimum amount of input data on the other facilities in order to provide reasonable estimates of impedance levels at these facilities. The impedances may be used in large network simulation models, and the methodology itself may be extended to the evaluation of alternative functional arrangements within the terminal complex.

ACKNOWLEDGMENTS

Work on this project was done as part of a contract from the Office of High Speed Ground Transportation (OHSGT), U. S. Department of Transportation. The authors acknowledge the assistance of the contract officer and his staff at OHSGT, many persons within the various carriers and terminal operating authorities, and Mr. Peter Dyson of Peat, Marwick, Mitchell and Co., who was responsible for the data collection effort.

REFERENCES

1. Lee, A. M. Applied Queueing Theory. The MacMillan Company, New York, 1966.
2. Saaty, T. L. Elements of Queueing Theory, with Applications. McGraw-Hill Book Company, New York, 1961.
3. Traffic Engineering Handbook. Institute of Traffic Engineers, 1965.
4. MacDorman, L. C. An Investigation of Pedestrian Travel Speeds in the Business District of Washington, D. C. Catholic University, Washington, D. C., May 1967.
5. Molina, E. C. Application of the Theory of Probability to Telephone Trunking Problems. Bell System Tech. Jour., July 1927.
6. Paullin, R. L. Passenger Flow at the Departure Lounges. Institute of Transportation and Traffic Engineering, University of California, July 29, 1968.

An Application of Marginal Utility to Travel Mode Choice

GORDON A. SHUNK and RICHARD J. BOUCHARD, Alan M. Voorhees and Associates, Inc.

The increased interest in planning for transit as a portion of the solution to urban transportation problems has generated concern for the effectiveness of traditional modal-choice prediction procedures. The concern is centered on the ability of models calibrated on data reflecting base-year transportation service to predict the results of marked changes in service. A model that is basically more behavioristic in nature rather than simulative conceivably could be an approach to a solution of the problem. A modal-choice relationship was developed that utilized as its independent decision variable a composite of several, more traditional factors. It has been theorized that components of this variable represent the disutilities of travel by competing modes as perceived by the traveler. The differences in disutility represent the marginal disutility of a given mode. Marginal disutility was the decision variable for traveler choice between auto and transit. The final decision variable combined out-of-pocket cost of transit and highway travel, family income, and parking cost, as well as travel time for the trip. Modal choice was examined for 1958 data from St. Paul and Minneapolis, using marginal disutility as the independent variable. The results reproduced base-year transit travel patterns very well without using traditional curve-fitting calibration procedures. The distribution of the results appeared to approximate quite closely the normal. This seems to indicate that the variable used may well approach the actual, though unperceived, variable on which modal choices are based. The technique has the benefits of conceivably alleviating the need for calibration and offering a more basic, behavioristic formulation that may transcend constraints of transit service levels.

•THE INCREASING AWARENESS of the role of mass transit in the solution of the problems of urban transportation has stimulated new concern about conventional methods for forecasting transit travel patterns. A principal element of this concern has been associated with the need to reflect properly the effect on patronage of markedly improved levels of transit service. Such a need is particularly important when rapid transit service is to be introduced in an area previously served only by conventional bus operations. The planners for rapid transit in most major U. S. metropolitan areas will have to face such a situation. Furthermore, proper reflection of the effects of service is paramount to analyzing the impact on travel of new concepts in urban mass transportation. The essence of this need is to estimate properly the reaction to service levels not previously available in a particular urban area, if indeed they had been available anywhere. The implication in this is the need for a universal or basic behavioral representation, the theory of which transcends all physical systems and service levels.

In 1968, Alan M. Voorhees and Associates, Inc., undertook for the Twin Cities Area Metropolitan Transit Commission a study of the efficacy of several alternative rapid transit systems for serving the twin cities of Minneapolis and St. Paul, Minnesota (1).

Paper sponsored by Committee on Origin and Destination and presented at the 49th Annual Meeting.

A portion of that study required estimation of the ridership-generating potential of the candidate systems. Among the systems examined in detail were a conventional bus system operated to take maximum advantage of express highways; a modern rapid rail transit; and a small-vehicle, high-speed, "new concept" system. To estimate patronage for systems having such a broad range of service levels, it was necessary to develop a procedure that would properly reflect subtle service differentials among systems and between transit and highway alternatives presented the traveler. An additional factor required that the procedure have especially broad applicability.

The only detailed information available on travel in the Twin Cities was from a 1958 origin-destination survey conducted by the Twin Cities Area Transportation Study (TCATS). The bus and highway systems and travel of 10 years previous were therefore the only means available to calibrate the procedure.

The cornerstone of any such patronage-estimating procedure is a modal-split model. Such a model determines the split or allocation of all person trips among the available travel modes. Modal-split relationships usually use characteristics of the trip, the trip-maker, and the elements of the transportation system as independent variables. They also may employ any of several measures of transit service. Combining measures of these effects yields a reliable method for predicting transit patronage. Using data from an origin-destination survey and an inventory of the transportation system, one may observe the proportions of total trips made on transit by people of various socioeconomic groups and with varying conditions of transit service. These data are used to develop geometric or mathematical relations between important factors and travel mode choice. The relationships are generalized and used to estimate transit ridership in the base year. The estimate is compared with observed travel volumes, and the predictive procedure is adjusted to compensate for discrepancies. The process is repeated until the prediction is within tolerable limits of accuracy. Information on future transit service and socioeconomic and density factors is used as input to the model to predict patronage on proposed transit systems. The proposed systems are adjusted as necessary to achieve the best ridership possible within defined economic tolerances.

APPROACH

Modal-split relationships are of two generic types, predistribution or postdistribution. These types refer to whether the modal allocation is made prior to distributing trips between origins and destinations. Predistribution allocations generally tend to make primary use of the characteristics of the trip-makers and their residence zones. They cannot use the most effective measures of transit system service because these measures are related to zonal interchanges; i.e., they are effective on the trip itself. Predistribution allocations, therefore, must rely on weaker service measures. Such procedures are less difficult and less expensive than postdistribution types, and they are strongly wedded to existing and historic transit service levels.

Because the Twin Cities program was oriented to show the effects of greatly improved transit service, it was decided to use a postdistribution procedure. Postdistribution techniques allocate trips among the various modes on an interchange basis; i.e., the distribution of total trips from all origins among all destinations is assumed completed. Then, based on transit service levels for each zonal interchange (origin-destination pair) as well as other factors, the allocation of total travel is made among the available modes. Postdistribution procedures permit the best possible reflection of the effect of transit service differentials that exist between different trip interchanges. This is very important because not all trips from a zone will use transit at the same rate because all trips are not destined to places where transit provides good service.

The term "service level" has broad implications. It includes at least implicitly each of the following factors: first walk time to transit, first wait time for transit, transit speed, stop frequency of vehicles, fares, number of transfers required, wait time at transfer, and walk time from transit.

Service level usually is not examined on its own merits. Modal allocation of trips implies competition among modes and, therefore, examination of relative service

levels. Previous modal-split models have used access or travel time ratios of transit to highway to represent relative transit service. The Twin Cities work used time differences between highway and transit travel. Time differences now are thought by some to better reflect differential service than ratios, although the theoretical basis for ratios is strong (2).

It has been known for some time that access and other nonmoving time components of total door-to-door time are perceived by the traveler in a different manner from time spent in motion. This is probably attributable to psychological factors that come into play as a person sees he is making no progress toward his destination as he waits, but progress is obvious when the vehicle moves. Nonrunning time, including walking time, occasionally has been given more influence on the computation of total door-to-door time than the running time. This influence has been effected by weighting the nonrunning or "excess" time by an empirical factor, usually evaluated at about 2.5 (2, p. 15). This procedure was employed in the Twin Cities work.

It follows that reducing running time differential by providing high-speed transit vehicles would have less effect on patronage than would reducing excess time. Selection of new transit systems should be directed at reducing those elements of transit travel time that are most effective in enhancing the relative attractiveness of transit. Alternatively, new transit systems may offer markedly greater speeds and commensurately reduced running times to offset uncorrectable excess times. These two concepts in improved service are the essential elements of new transit systems currently being developed.

Modal choice also is influenced by characteristics of the trip-maker, his trip, and the locations at which the trip starts and ends. Trip-maker characteristics that most influence modal choice are sex, age, and income. These factors occasionally are represented by a surrogate population density of the origin zone. Income is highly correlated with car ownership, which has a direct effect on modal choice. Parking cost has a very significant effect on travel mode selection; it may be represented by the surrogate employment density. The Twin Cities modal-split relationship utilized zonal, median family income to represent the effect of trip-maker and residence zone characteristics. Parking cost was used to represent destination zone characteristic effects.

Trip purpose has a great effect on modal choice. Work trips are usually the most oriented to transit. This may be attributed to such things as regularity, occurrence during hours of dense travel, and most employment historically being located in the central business district (CBD), the area best served by transit. School travel is also transit-oriented, whether the transit is public or school bus. People are less inclined to rely on transit for trips with other purposes, probably because the destinations for these trips are usually in less densely developed areas that, consequently, are not well served by transit. These trips are also less regular, and the timing of them is less important, so they can be made when the family automobile is available. Transit trips for these purposes are dominated heavily by transit captives, persons who have no automobile available permanently or temporarily. The modal-split relationships discussed here were for work trips only.

Most modal-split or allocation procedures employ and are referred to as "models." This terminology is consistent because some mathematical formulation is used. The Twin Cities relationships were not developed as mathematical relationships but as graphic plots of nonlinear relationships. To develop these, it was necessary to employ manual plotting of stratified observations. This approach was chosen because current, nonlinear, regression techniques are constrained by limits on order and requirements of consistency. Data describing the trip-maker, trip, transit service, and highway service were cross-stratified in several ways by a specially tailored computer routine. The transit percent of total ridership observed in each cell of the cross-stratification was plotted against the several strata levels. The curves or surfaces describing variation in transit ridership thus were defined. Manual plotting permitted use of multiple-dimension, curvilinear relationships. It did not require assumptions or constraint of data to force linearity. It provided the additional advantage of allowing extrapolation of the curve in regions of interest for which data were not available in a manner dictated

by experience. Because the objective of this study was to analyze the effects of service that represents a major improvement over that currently available, the extrapolation permitted developing more complete and consistent curves.

THE CONCEPT

Development of the Twin Cities modal-split relationship involved examination of a universal utility measure. Many people have thought for some time that modal choice is much akin to traffic diversion on highways, albeit between different means of travel rather than different routes (3). People either choose or are diverted to alternative travel modes by their perception of the relative attractiveness of each. The important element of this theory is a hypothetical factor to which all others can and must be reduced. This is the element on which people at least implicitly may base modal choices. This factor, termed "modal-choice utility" in this study, may be represented properly by a combination of several, more basic factors of influence such as time and dollar cost. The Twin Cities work undertook examination of the essence and applicability of this hypothesis. Such an approach could lead to much more easily applied modal-choice relationships because of the implied universality of the utility factor. It also would permit inclusion of such economic factors as road pricing and other economic policies. Additionally, it would validate extrapolation in areas where calibrating data were sparse by virtue of the theory of the utility function.

The theory of the utility function is of this nature: Given that a variable can be defined that explicitly or implicitly represents the datum on which people base decisions, the distribution of the results of such decisions plotted against values of the variable will approach normality. Thus, if a decision variable represents all perceived travel disutility, it should be a candidate for such a function. If the results of observations of a dependent variable appear to be distributed normally when plotted or examined, it may be assumed that the decision variable is adequate. The following section will elaborate on the testing to prove such a hypothesis.

PROCEDURE

The procedure for developing a modal-split relationship, regardless of its formulation, follows a basic pattern. This pattern, in general, consists of matching existing or observed ridership to corresponding socioeconomic characteristics of the trip-maker and service characteristics of the transit system. A relationship is developed, tested, revised to match observed conditions, and retested. The process is iterated to satisfactory closure tolerances. The final relationship is applied to future-year person travel estimates using future-year socioeconomic and transit system characteristics as independent variables. The independent characteristics used in development of the relationship must have been predicted for the future year.

Development of the Twin Cities modal-split relationship began with preparation of survey-year data. Major effort was concentrated on preparing a representation of the 1958 transit system for computer processing. This effort included coding transit routes and their characteristics for input to the Department of Housing and Urban Development (HUD) Transit Planning Program package. This interrelated set of programs is capable of representing most aspects of a transit system that are important to modal choice and operational analysis. It also permits development of data that are technically and physically consistent with data currently developed for highway systems using the BPR BELMN package. The two sets of data then can be compared. The 1958 Twin Cities highway network was prepared using BELMN and inputs provided by the Minnesota Highway Department. These data were the same as those used by the department in model calibration for the TCATS. The HUD transit programs permit coding of a transit system and gaining access to it in such a manner as to have available for individual analysis such components of transit service as walk, wait, transfer, and run time; number of transfers; and fare. Coding of the 1958 Twin Cities transit system was done from route schedules, thereby using schedule stops, times, headways, and fares.

Travel observations were based on the internal trip cards from the TCATS 1958 origin-destination survey. Several characteristics of the trip interchange represented by each trip card then were appended to individual trip records. These included all time components of the transit and highway trips, the transit fare, and the highway distance. Other factors included were median income of the production and parking costs of the attraction zones.

The time and distance components of each trip interchange could be combined in various ways to test the effect of alternatively constituted independent variables on modal choice. These data, describing the character of the trip, travel service, and origin and destination zones, were used as bases for multiple cross-stratification summaries. Such summaries permitted examination of the variation in observed 1958 transit ridership between cells defined by each chosen level of each independent variable. The cell modal-split values were plotted against the principal independent variable (usually a time- and cost-related function) in the number of dimensions dictated by the chosen cross-stratification. This procedure permitted analysis of the effects of certain variables at different levels of others. It also allowed a crude analysis of the separate effects of usually dependent factors.

The various combinations of separate variables represented utility measures. In such utility variables, all measurable costs of travel, both time and money, were combined for each mode separately in a function with a common unit of measurement. This approximated the theoretical disutility of travel for the given mode. The disutility by each mode was computed, and the difference between them was found. This differential disutility represented the marginal utility of one mode over another. The times used in the disutility computation were equivalent times, which include all elements of travel time: walking, waiting, and riding. Riding time is that time spent in a vehicle, and it includes intermediate stop time by the vehicle. Excess time is the remaining time spent in traveling from door to door. For transit, excess time includes walking and waiting times; for automobiles, it includes walking and parking times. Transit waiting times are evaluated as half the headway (interarrival time) of the transit line chosen. The equivalent time was computed by combining riding time with excess time, the latter weighted (multiplied) by 2.5. The variables in the model and their factors are given in Table 1.

The cost for transit was the fare; the cost for automobile was parking plus out-of-pocket operation cost, i.e., gasoline, oil, and maintenance costs and their associated taxes. The operation cost was directly related to trip mileages; a unit value was multiplied by trip distance for each particular trip interchange to obtain trip travel cost. The cost per mile was based on a figure developed by the Bureau of Public Roads. This figure was modified to reflect speed differences between non-CBD- and CBD-oriented trips. Parking costs were halved to effect allocation between the two portions of attracted trips. All costs were converted to time units by dividing them by a variable cost of time, computed as a percentage of income. Based on the results of other studies, the time cost was evaluated at 25 percent of income (4). The resulting cost of the trip, in time units, was added to the equivalent trip time for the respective mode; and the difference between these travel disutilities was computed.

Percent transit (modal split) was plotted graphically against this marginal disutility. These plots were made for three stratifications of zonal median family income: less than $6,000; $6,000 to $8,100; and over $8,100. These strata

TABLE 1

VARIABLES AND WEIGHTING FACTORS IN THE TWIN CITIES MARGINAL UTILITY MODAL-CHOICE MODEL

Variable	Symbol	Factor
Walk time to/from transit	T_a	2.5
Wait time for transit	T_w	2.5
Transit running time	T_r	1.0
Transit fare	F	1.0
Auto terminal time	A_t	2.5
Auto running time	A_r	1.0
Parking cost	P	0.5
Highway distance	D	4.0, 5.7[a]
Marginal utility[b]	U	—
Cost of time[c]	C	—

[a]These are cost-per-mile factors rather than weights. For trips attracted to CBD, 5.7 cents per mile was used. Other trips cost 4.0 cents per mile.

[b]Computation equation for marginal utility of auto over transit for non-CBD trips:

$$U = 2.5(T_a + T_w - A_t) + (T_r - A_r) + (F - 0.5P - 4.0D)/C$$

[c]Cost of time is computed as 25 percent of income:

$$\left[\frac{\text{Annual income (cents/year)}}{\text{(2,080 hours/year) (60 minutes/hour)}}\right] \times 0.25 = C$$

were established by examination of the frequency of occurrence in several income groups. The three groups chosen contained nearly equal numbers of trips. Three groups were selected because this was the largest number that yielded adequate observations in the classes. The results of these plots, shown in Figure 1, were very consistent and appeared to follow the theory regarding normal distribution quite well.

A plot of modal choice against marginal disutility was made on log-probability paper to test the theory. In this plot (Fig. 2) observations were not stratified by income as in the actual model. The conformance with the straight line shown implies that the basic mathematical function, which is hypothesized as being a normal distribution, is of an exponential nature. It is interesting to note the dispersion and limits of observations at the tails of the curve. Pratt (3) has suggested that this results from the captivity of people to various modes; i.e., free choice does not occur because of other constraints. There is, therefore, a tendency for observations to flatten in these regions, not penetrating some artificial limits. This tendency results in the aggregate from certain proportions of people using transit or auto regardless of the relative utility of either mode. In turn, these proportions are attributable respectively to people who cannot use an automobile or who must have an auto at work regardless of how relatively good transit service is.

The three-curve utility model then was coded for a specialized computer program. This program read the 1958 work person (all modes) trips, used the model to determine which of those trips would use transit, and compared the transit trip prediction to the observed 1958 transit trips. The comparisons were made in several ways: (a) by trips produced, (b) by trips attracted, (c) by trip interchange, (d) by income group, and (e) by marginal disutility level. It thus was possible to determine quite definitively how well the model was functioning at all significant levels. Trip-length distributions and other graphic displays were used to further examine model results.

RESULTS

The first run of the work trip, modal-split model calibration indicated that trips to the CBD areas were being somewhat underpredicted. The result appeared to be attributable to inadequate parking cost figures. The figures used initially were for traffic analysis districts rather than zones. It was apparent that aggregation had masked some of the true influence of parking cost on transit patronage. A relationship was developed describing the variation in parking costs with employment density. This relationship was based on aggregate district data and appeared quite consistent with similar curves developed for other areas. Using this curve and zonal employment density figures, new zonal parking costs were developed. The model was run again with the zonal parking costs, and the results were considered acceptable.

Statistics representing the prediction accuracy of the work modal-split model are shown below. These are comparisons between parameters of the observed and estimated transit trip tables for 1958.

R^2, transit productions by zone	=	0.880
R^2, transit attractions by zone	=	0.965
R^2, transit productions by district	=	0.999
R^2, transit attractions by district	=	0.999
R^2, transit interchanges by zone	=	0.571
Average airline transit trip length	=	+2.6%
Average transit trip time	=	-0.3%
CBD transit trips	=	-1.8%
Total transit trips	=	-0.2%

The percents of variability explained (R^2) by the model are shown by zone for productions, attractions, and interchanges and by district for productions and attractions. District transit productions and attractions were predicted accurately in almost very case. Scatter plots of the observed and estimated productions and attractions by district are shown in Figures 3 and 4. Zonal attractions, the focus or high-density confluence of

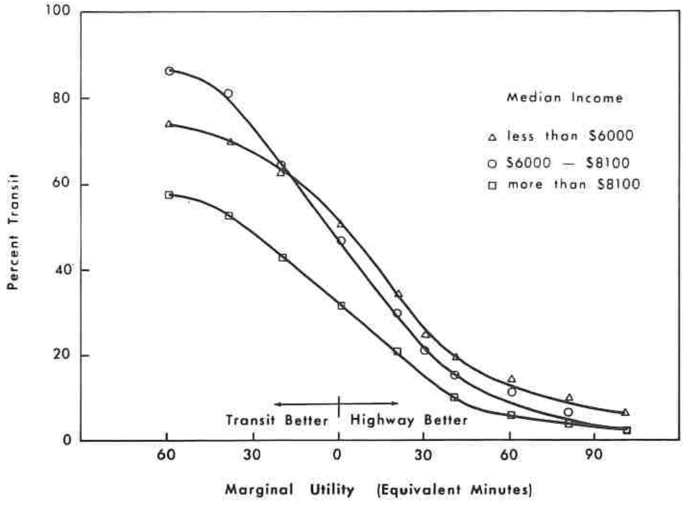


Figure 1. Work trips in Twin Cities, marginal utility, modal-choice model.

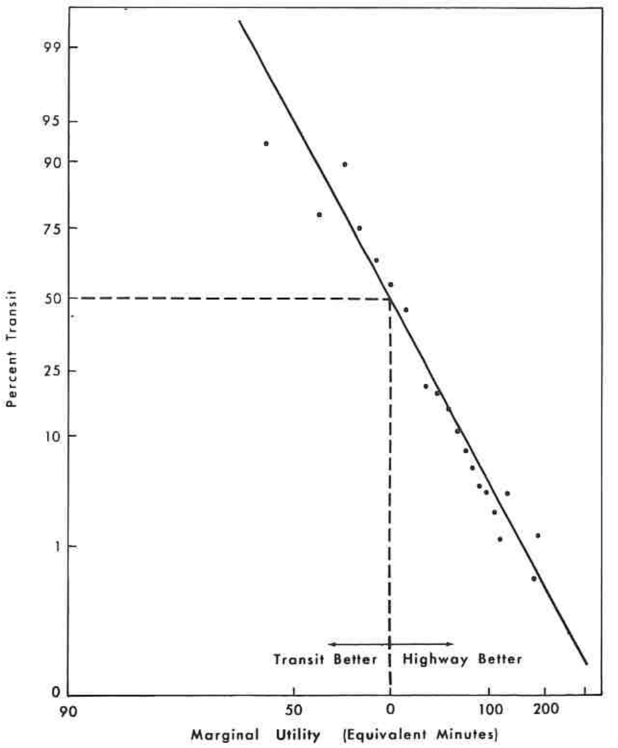


Figure 2. Log-probability relationship; percent transit versus marginal utility work trips.

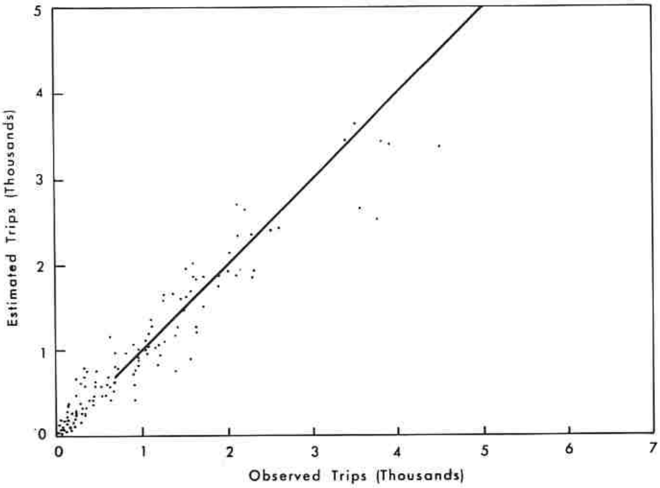


Figure 3. Transit work trips, Twin Cities, 1958 (observed and estimated productions by district).

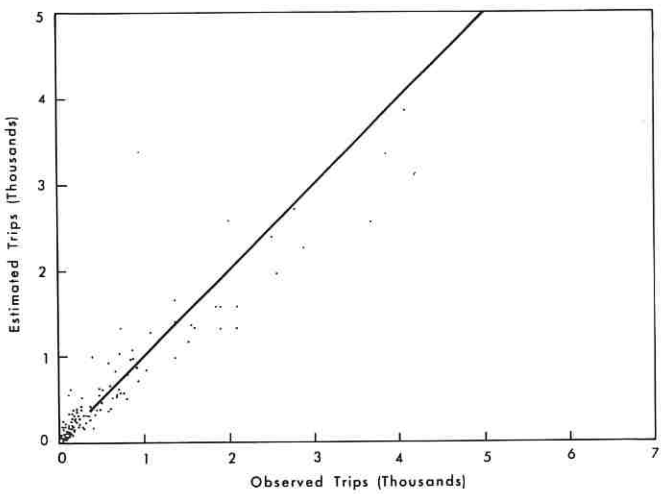


Figure 4. Transit work trips, Twin Cities, 1958 (observed and estimated attractions by district).

work trips, were predicted extremely well. Even production of transit trips was predicted very well. The attraction results imply that parking cost is indeed a good indicator of the attractiveness of a zone for transit travel. The lower R^2 value for productions implies that the income variable used is less reflective perhaps of transit trip production ability than of something else, such as auto ownership. The apparently low evaluation of transit interchange prediction performance is actually quite good. Statistics for the same comparison run considerably lower in other studies. The impact of a good result here is that actual travel movements or patterns were predicted accurately 57 percent of the time. This reflects the validity of the interchange service

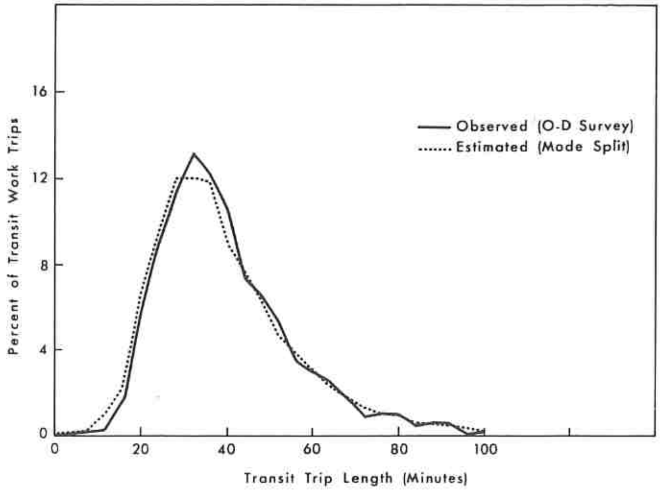


Figure 5. Total transit time distribution in transit work trips, Twin Cities, 1958.

variable—marginal utility. Prediction of the number of transit trips in total and to the CBD (the major generator of transit travel) are of obvious importance. These model results were again excellent.

Details regarding transit trip length too often are not compared. Such a comparison is important, however, especially for cost analysis purposes. The implication involves person and vehicle-miles of travel on the transit system and their impact on revenue and operating costs. Two analyses of this aspect were made. First, a comparison of trip length, as measured by total transit trip time, was made between the observed and

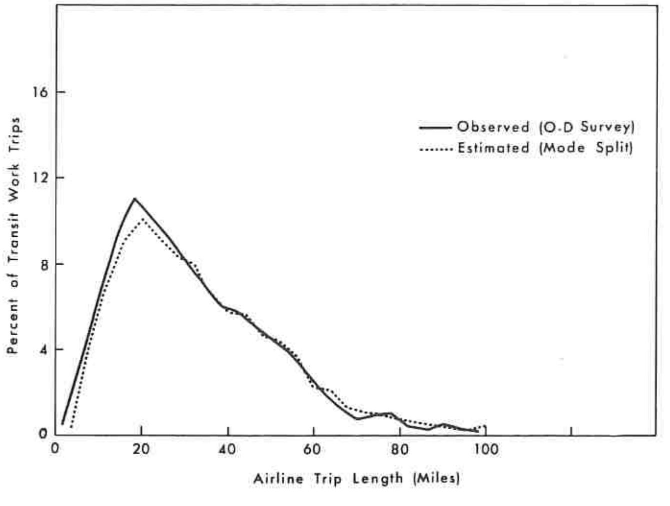


Figure 6. Airline distance distribution in transit work trips, Twin Cities, 1958.

estimated trip tables. Figure 5 shows that the average length of the predicted trips was within 0.3 percent of the observed trips. The distribution verifies this accuracy. It can be assumed then that forecasts using the model will represent accurately the extent of transit travel in the region. Second, observed and estimated airline trip lengths were compared. Such a comparison is valuable because it is not transit system-dependent and should remain valid if system service type changes radically. The distribution for this comparison is shown in Figure 6. Comparison of the average trip length in this case showed the model results within 3 percent of the observed results. The results indicate that the model survives tests of quantity and orientation.

It appears that this new approach to modal-choice prediction is capable of accurately representing an established condition. The results given were established with minimum use of traditional calibration. The parking cost adjustment was a revision of input data to improve accuracy. Several alternative model forms and parameter values were tested, but no subjective curve adjustment was employed. In short, the model appears to have the properties necessary for universal application. It may thus be possible, after a few refinements, to use such models without calibration. This would alleviate the need to have base-year data for calibration. Further tests of the technique are being executed and proposed at this time.

ACKNOWLEDGMENTS

The efforts, advice, and guidance of the following people in the work reported in this paper are acknowledged most gratefully: Thomas B. Deen, Richard H. Pratt, Marvin Golenberg, Gordon W. Schultz, and Francis P. D. Navin.

REFERENCES

1. Proposed Work Program for Technical Studies for Transit Improvement in the Twin Cities Area. Prepared for the Metropolitan Transit Commission of the Twin Cities Area by Alan M. Voorhees and Associates, Inc., McLean, Va., 1967.
2. Quarmby, D. A. Transport Planning and the Choice of Travel Mode. Summary Report to the Transport Committee of Leeds City Council, Leeds, England, 1966.
3. Pratt, R. H. A Utilitarian Theory of Travel Mode Choice. Paper presented at the 49th Annual Meeting and included in this RECORD.
4. Lisco, T. E. The Value of Commuters' Travel Time—A Study in Urban Transportation. Univ. of Chicago, PhD dissertation, 1967.

A Utilitarian Theory of Travel Mode Choice

RICHARD H. PRATT, R. H. Pratt Associates, Garrett Park, Maryland

Intensified study of public transit service has placed demands on the transportation planner for improved choice of travel mode modeling techniques. An underlying theory of modal choice is needed to correct inadequacies in our present approach. A theory is proposed that is built on the suppositions that individual choice of mode is utilitarian, that common measures of individual trip utility are subject to chance errors describable by the normal distribution error function, and that deviations result from predictable influences. A difference in disutility measure is set forth for comparison of travel utility. The measure combines time, convenience, and dollar cost into a common unit of equivalent time. The probability of free choice of a given mode is described mathematically as a function of the possible disutility savings. The formulation is postulated to be the normal probability density function, predicting 50 percent probability at zero disutility difference. Submodal-split study results pertaining to both transit and highway route choice are examined and found to support the free modal choice mathematical description. Deviations to be expected in applying the theory to choice of prime mode are examined. It is assumed that long-term captivity to transit or auto can be expressed as a constant probability, and the resultant constrained formulations are illustrated. Effects of excessive trip length and desirable operational refinements are discussed. It is concluded that the proposed theory is readily applicable to modal-choice forecasting and multimode analysis and may lead also to broader applications.

•THE CURRENT INTEREST in providing significant public transit service as part of the total transportation system shows no likelihood of diminishing in this era of concern with urban needs. Demands on transportation planners for improved evaluation of the interplay between private auto and public transit usage can thus be expected to continue and grow.

Predictive models for forecasting choice of travel mode have undergone extensive improvement in the past decade. Nevertheless, travel analysis still suffers from lack of a generally accepted underlying theory of modal choice. Present operational models are mostly individually tailored, empirical formulas or hand-drawn experience curves.

THE ADVANTAGES OF A THEORY

A satisfactory theoretical explanation of observed modal-choice behavior would provide benefits in forecasting travel and in understanding user evaluation of transportation system attributes. A theory is needed to produce advantages such as the following.

1. A modal-choice technique based on a satisfactory theory would allow use of a pretested model requiring only local calibration using standardized procedures. This would put less strain on available survey sample sizes and save time and talent now expended on developing individually tailored models for each application.
2. A proven theory would provide a sound basis for extrapolation beyond those sets of time, convenience, and cost alternatives presently observed in the environment of conventional transportation systems. In contrast, the validity of an empirical modal-

Paper sponsored by Committee on Origin and Destination and presented at the 49th Annual Meeting.

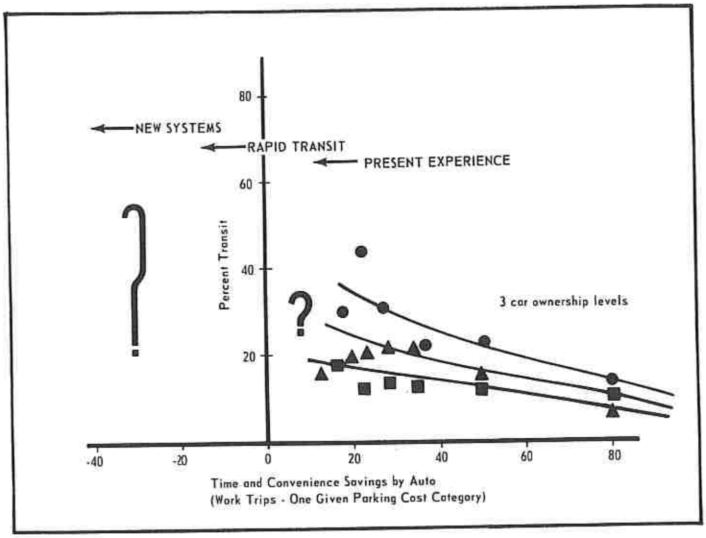


Figure 1. The empirical curve extrapolation problem.

split model under conditions not found in the existing city is dependent in large measure on the experience and judgment of the analyst and on his skill in extrapolating the predictive relationships. Figure 1 shows the difficulties encountered with such present-day models.

3. A model structured on theory would provide a reference point for measurement of travel preferences and thus would broaden the understanding of what travel characteristics are considered important bv transportation consumers and how they are weighted. This in turn would provide more accurate input into cost-benefit analyses and would allow identification of those transportation system attributes that deserve priority attention in development and design.

4. A modal-split model with a theoretical framework can be more readily understood, defended, and subjected to critical examination. An empirical model must be judged primarily on the basis of consistency of results. A model based on theory can be evaluated both on this basis and by examining the inherent logic or experience with the postulates expressing user behavior and preference from which the model has been derived. The technical strengths of the model can be identified and used to their best advantage; the weaknesses of the model can be isolated and made the subject of further research.

THE BASIC PROPOSITION

The choice of mode theory presented here for consideration is comprised of primary suppositions and corollary statements as follows.

Supposition 1—Individual choice of mode is utilitarian. The trip-maker's concern is to minimize the sum total of personal disutility involved in the travel action. Corollary: If an individual's unique perception of the sum disutility of using each of the alternative travel modes could be measured for a given trip, his choice of mode could be absolutely predicted.

Supposition 2—Description of individual utility perceptions with standard network analysis techniques is affected by a multitude of chance errors. These include discrepancies caused by the variations in behavior related to differing individual value systems. Corollary: Individual choice of mode cannot be forecast; but as a substitute, the probability of an individual choice can be predicted using the error function associated with the normal distribution.

Supposition 3—Deviation from the normal distribution can be explained by such predictable influences as captive riding and resistance to long trip length.

The corollary of the second supposition is simply a restatement of the central limit theorem of probability mathematics, provided one agrees with the related supposition that the errors involved are errors of chance. The following is a partial list of discrepancies and errors that may occur in measuring individual perception of travel utility.

1. Errors inherent in averaging, specifically the use of one set of travel parameters to represent conditions facing all individuals in a given trip interchange population;
2. Discrepencies caused by variations among individuals in the perception of the utility parameters, in other words, variations caused by individual responses based on imperfect information;
3. Discrepancies caused by behavioral variations among individuals in the evaluation of utility;
4. Random network measurement errors in determining and using the mean travel parameters of alternative modes available to the trip interchange population; and
5. Network biases and errors in specification of the utility measure.

It seems quite reasonable to classify the first four categories as comprising errors of chance, satisfying the proposition in this regard. The fifth type is comprised of consistent, nonrandom errors and must be eliminated insofar as possible in any successful application of the theory.

A COMMON UTILITY MEASURE

For testing and applying the postulated theory, it must be possible to have a realistic and common measure of trip utility. If trip-generation rates are held constant, benefits accruing from reaching the trip destination are irrelevant; and modal-choice evaluation becomes simply a function of the relative opportunity to reduce travel time, inconvenience, and cost. An appropriate and apparently satisfactory measure is "difference in trip disutility," already used by D. A. Quarmby (1) and others in travel forecasting.

In this common measure, all trip costs, including but not necessarily limited to travel time, inconvenience, and money, are converted to equivalent values. They are summed for each traffic interchange and mode under consideration. The comparison of alternative modes is then made for each interchange on the basis of the algebraic difference in their respective disutilities.

In the theory being presented here, the probability of an inferior mode being used is described as a function of misclassification of individual utility perceptions for individuals within a population. This requires knowing the difference between the two measured values for the alternative modes; thus ratios cannot be considered properly for use. Methods for developing the equivalence values used in constructing the disutility difference measure are secondary to the choice of mode theory itself and are discussed in reference to applications.

FREE CHOICE MATHEMATICAL DESCRIPTION

Using the "difference in disutility" measure, the mathematical relationships implicit in the first and second suppositions of the postulated theory can be derived. The logic is outlined using a hypothetical comparison of travel mode B with alternate mode A as shown in Figure 2. Free choice of mode is assumed.

In accordance with the theory, individually perceived disutility differences will occur at variance with the value as measured by the traffic analyst. This is represented in the upper part of Figure 2 by a normal error distribution drawn around each of five measured disutility differences. Now, in line with the supposition of utilitarian choice, a trip-maker will choose mode B if the difference is positive in favor of B. Some individuals, because of the variance that has been described, will perceive the disutility saving to be positive even though it has been measured otherwise. These individuals are misclassified.

The misclassified individuals are a function of the area under the normal curve where the disutility difference is of opposite sign from the measured value. It follows that the probability of a trip-maker choosing the mode that has been measured as being inferior is equal to the probability of misclassification. Following this line of reasoning, the probability of using mode B can be plotted for different disutility differences (bottom of Fig. 2). The curve obtained is mathematically described by the normal probability density function.

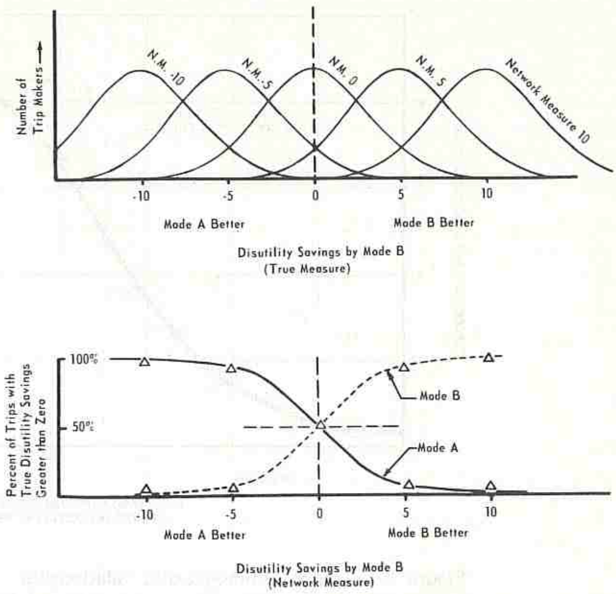


Figure 2. Free choice of mode mathematical relationships.

The most pertinent elements of the utilitarian-theory mathematical description as it pertains to the free choice of mode can be summarized as follows.

1. The probability of free choice of a given travel mode is a function of the disutility savings obtained through use of that mode as compared to the alternate.
2. The probability is described by the normal probability density function.
3. The resultant predictive curve has its point of inflection at 50-percent probability and zero measured disutility savings.

When travel captive to a particular mode is considered, the mathematical description must be modified. The effect of captive travel will be discussed subsequently as will the effect of excessive trip length.

EXAMPLES FROM SUBMODAL SPLIT

Submodal-split traffic analyses provide a specialized source of data appropriate for testing the first two suppositions of the proposed theory before proceeding to describe the likely form that deviations from the normal distribution may take when captivity and resistance to long trip length are involved. Submodal split applies to the special case where the trip-maker has already been assigned to the transit or auto mode. The remaining question to be answered is whether he will choose bus or rail routing if he is a transit rider, or freeway or arterial routing if he is an auto user. Obviously, no transit rider is captive to rail or bus if both services are available to the public. Neither is any auto driver captive to using either freeways or arterial streets. All decisions are free choice and the normal probability density function should be found to hold without deviation.

The first example is provided by the derivation of a rapid transit versus surface transit diversion curve as discussed in a paper by the author and Thomas B. Deen (2). The data used were from travel surveys covering the Chicago Transit Authority's "Skokie Swift" rapid transit operation and paralleling bus routes.

In the case studied there was no fare differential between transit submodes. Therefore, difference in disutility was expressed using only time and convenience measures, summed and designated as "equivalent time." Convenience was quantified in terms of "excess time" and was comprised of the walking, waiting, and transfer time involved in any door-to-door trip. Excess time was weighted by a factor to

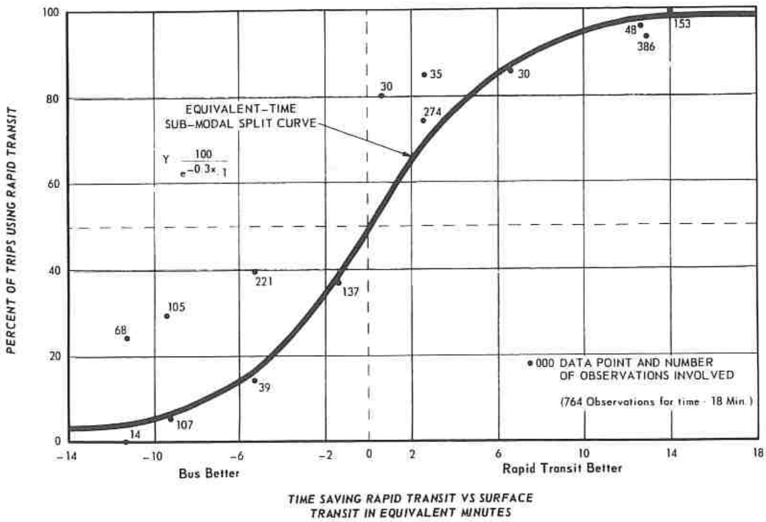


Figure 3. Transit submodal-split relationship. (From Richard H. Pratt and Thomas B. Deen. Estimation of Sub-Modal Split Within the Transit Mode. Highway Research Record 205, 1967, pp. 20-30.)

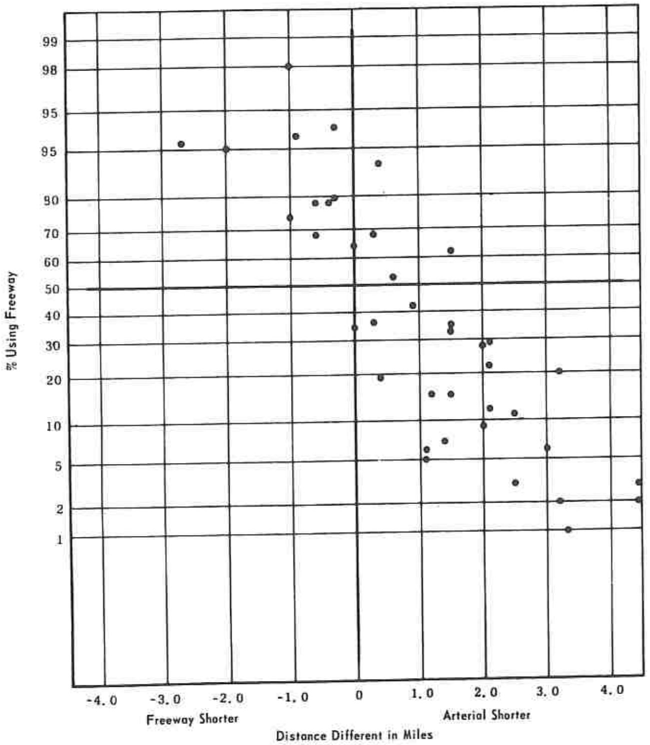


Figure 4. Freeway usage related to distance difference. (From Howard W. Bevis. Estimating a Road User Cost Function From Diversion Curve Data. Highway Research Recrod 100, 1965, pp. 47-54.)

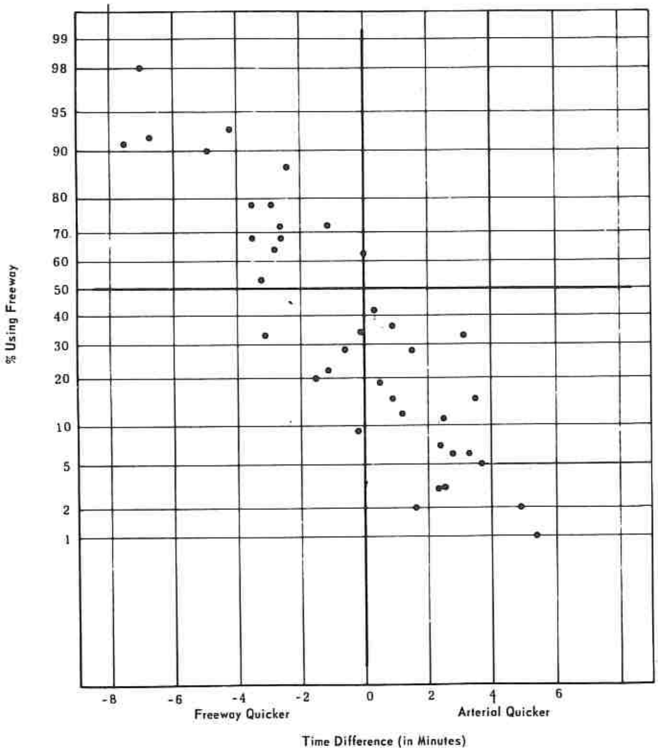


Figure 5. Freeway usage related to time difference. (See Fig. 4 for source.)

render it psychologically equivalent to running time. Several factors were tested and the value 2.5 was chosen.

The percentage of transit trips observed using rapid transit was plotted against the corresponding equivalent time saving (Fig. 3). The R^2 of the S-shaped curve that fitted, comparing predicted and actual percentage of submodal split on an interchange basis and weighting by the number of observations, was 0.886. The formulation, as illustrated, was a logistics curve with the point of inflection at 50 percent and zero difference. For purposes of this discussion, it can be considered an acceptable approximation of the normal distribution function. Thus, support is provided for the proposed theory.

A second submodal-split test of the utilitarian theory is provided by the work of Howard W. Bevis (3) on road-user cost functions. Bevis theorized that auto driver choice of route could be described through use of the normal probability density function if one compared routes in terms of generalized cost. This is an equally acceptable way of expressing difference in disutility, the common unit of measure simply being cents instead of minutes.

Bevis analyzed freeway diversion curve data for freeways in Washington, Dallas, Houston, and San Diego. Figures 4 and 5 show the percentage of freeway trips plotted on probability paper against distance difference and time difference respectively. The linearity of the plots in both instances provides support for the assumption of normality.

Combining the distance and time components into a generalized user cost, Bevis formulated equations constrained to reflect 50-percent diversion at zero cost difference. Correlation coefficients in excess of 0.80 were obtained for each of the data sets tested. Bevis concluded that the data fit the constrained normal cumulative distribution function very well.

APPLICATION TO PRIMARY CHOICE OF MODE

Examination of the utilitarian theory of modal choice using submodal-split data gives every indication that the normal cumulative distribution holds without any significant deviation when only free choice is involved. Using an assumed variance, the generalized predictive curve on standard linear coordinates is shown in Figure 6-A.

Turning to the prime choice of mode, auto versus transit, it is reasonable to expect certain specifiable deviations. In particular, captivity to either mode will restrict the portion of the trip-making population having free choice and should correspondingly affect the shape of the predictive curve.

If it can be assumed that a given range of average incomes defines some constant probability of transit captivity, then the free choice riding can be segregated out and investigated for applicability to the basic theory. Figure 6-B shows a hypothetical low-income population having a 20-percent constant probability of transit captivity. The remaining 80 percent of the population is assigned to free choice probability space and is allocated using the normal cumulative distribution. The theory will be tested using this type of probability space allocation.

It could be argued in opposition to the constant probability assumption that, even for a given income group, auto ownership should decrease as transit service improves, with transit captivity varying accordingly. This argument, which is valid in the usual sense of transit captivity, can be met through use of a more rigorous definition. Transit riders will be considered captive only if long-term unavailability of auto transportation is involved. The auto unavailability must be primarily independent of the quality of transit service available for the trip under consideration.

A transit rider will be considered captive only if the following criteria are met.

1. He does not have an auto available for the trip.
2. He cannot afford to buy and operate an auto.
3. He cannot find a ride or afford a taxi.

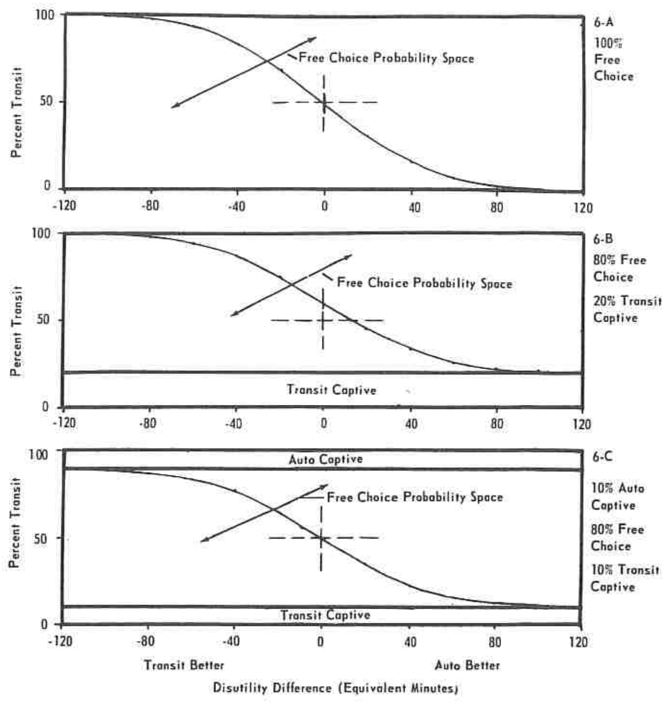


Figure 6. Captive and free choice probability space allocation.

With these criteria, a transit rider who can afford a car, even if he has none, will always be classified as a choice rider.

Auto captivity, too, must be considered. An auto user will be defined as captive if either of the following conditions are met:

1. He requires an auto at his destination (such as for work); or
2. He requires an auto on one leg of his trip (such as to make an intermediate stop on the way home).

It is important to note that the auto captivity definition, like its transit counterpart, is designed to be independent of the quality of transit service for the trip under consideration.

Figure 6-C shows a hypothetical moderate-income population with 10 percent constant probabilities of both transit captivity and auto captivity. Again, the remaining noncaptive population is assigned to a free choice probability space and is allocated as before.

The normal cumulative distribution as shown in Figure 6-A is a straight line on normal probability paper (Fig. 7). Also shown in Figure 7 are the hypothetical curves of Figures 6-B and 6-C replotted. They exhibit a curvature on normal probability paper that is introduced by the assumptions of constant captivity. The hypothetical plots of Figure 7 are compared with observed data in the next section.

EXAMPLES FROM PRIMARY MODAL SPLIT

A few prime modal-split analyses have been prepared with trip interchange comparisons between auto and transit service expressed in a manner approximating the difference in disutility measure. There is enough information, however, to allow tentative evaluation of the proposed theory. All data presented here are for work purpose trips.

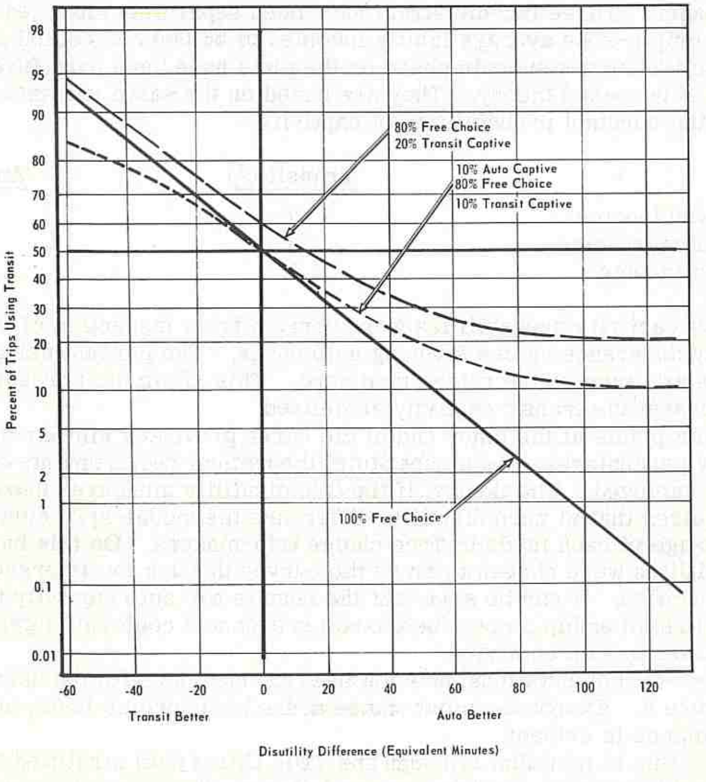


Figure 7. Captivity effects plotted on probability paper.

The first example is from the Twin Cities area of Minneapolis-St. Paul and uses input developed for modal-split model calibration by Alan M. Voorhees and Associates (4, 5). The trip data base was a 5 percent home interview sample taken in 1958.

Trip disutility measures were constructed from the following travel impedance components.

Auto	Transit
Driving time	Running time
Parking time	Walking time
Walking time	Waiting time
Auto operating cost	Transferring time
Parking charge	Transit fare

Time measures were estimated using highway and transit networks. Parking charges were split between the going and return auto trips. Auto-related dollar costs were not divided by auto occupancy on the assumption that cost savings obtained by carrying passengers are counterbalanced by the time and inconvenience involved in passenger pick-up and delivery.

Time and dollar quantities were converted to a common disutility measure using a series of equivalence factors. These factors were for expediency based on experience and prior studies by others. Driving and running time was used as the basic unit. Excess times, including auto terminal time, were factored by 2.5 as an expression of inconvenience. Dollar costs were converted to time equivalents by valuing time at 25 percent of the wage rate implied by the applicable zone-of-origin average-income estimate.

The results of comparing the percent of work trips using transit with the difference in disutility between auto and transit service are reproduced in Figure 8 as plotted on probability paper. Three income strata have been separately analyzed using as the break points origin-zone average family incomes of $6,000 and $8,100 per year.

The predictive curves superimposed on the plots have been hand fitted in conformance with the proposed theory. They are based on the same variance throughout and on the following constant probabilities of captivity.

	Transit (%)	Auto (%)
Lower income	7	7
Medium income	3	7
High income	1	25

The transit captivity probabilities were derived from inspection of the data points at high disutility difference values favoring auto usage. The percent using transit appears to approach a minimum value rather than zero. This minimum percentage has been taken to represent the transit captivity as defined.

Lack of data points at the other end of the curve prevented similar determination of auto captivity percentages. As a substitute, theoretical requirements of the proposed theory were employed. The theory, if the trip disutility measures used are accepted as valid, requires that at zero disutility difference the modal-split curve must predict 50 percent usage of each mode by free choice trip-makers. On this basis, auto captivity probabilities were chosen to force the curves through the intersection of coordinates thus specified. It can be seen that the results are auto captivity percentages that fall in logical relationship among the income groups and conform in general with available information on auto captivity.

Goodness-of-fit computations have not been carried out with the data in the format shown in Figure 8. Except for minor noise at the high-income level, however, very close conformance is evident.

It is interesting to note that although the Twin Cities data stratified by income provide considerable support for the proposed theory, the same data stratified by auto ownership produce noticeably skewed data plots. It can be concluded that whereas

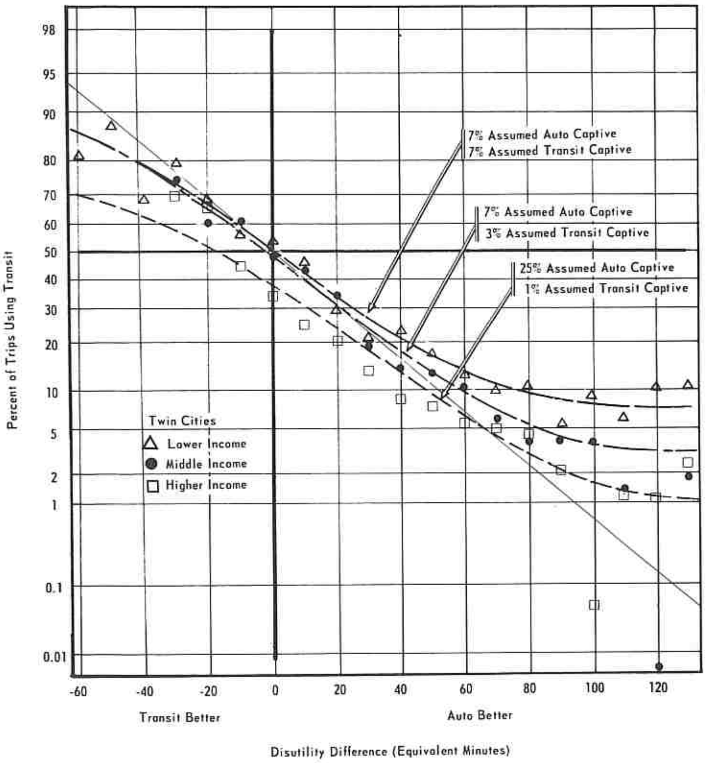


Figure 8. Twin Cities modal-choice relationships.

income provides a relatively unbiased basis for stratification, auto ownership levels do not. Income cannot be directly influenced by choice of mode; car ownership logically can be.

The second example using primary modal-split data is derived from 1955 Washington, D. C., home interview survey, work-trip information prepared by J. Royce Ginn (6). Ginn reprocessed the original Traffic Research Corporation modal-split model inputs (7), classifying trip interchanges by income rank and difference in equivalent time. The equivalent time measure incorporated the same travel cost components as were used for difference in disutility in the more recent Twin Cities studies except that auto-related costs were divided by auto occupancy. Most of the trip interchanges involved downtown-destined trips.

Ginn experimented with various equivalent time measures using a wide range of time equivalents. The data sets selected for presentation here used equivalents as follows: 2.00 equivalent minutes per excess time minute, 1.50 equivalent minutes per penny cost (lower incomes), and 0.75 equivalent minute per penny cost (upper incomes). These equivalents are in the same order of magnitude as the Twin Cities factors.

Figure 9 shows the Washington data in identical form as the Twin Cities data of Figure 8. The hand-fitted curves are based on the following constant probabilities of captivity:

	Transit (%)	Auto (%)
Lower income	18	3
Higher income	6	10

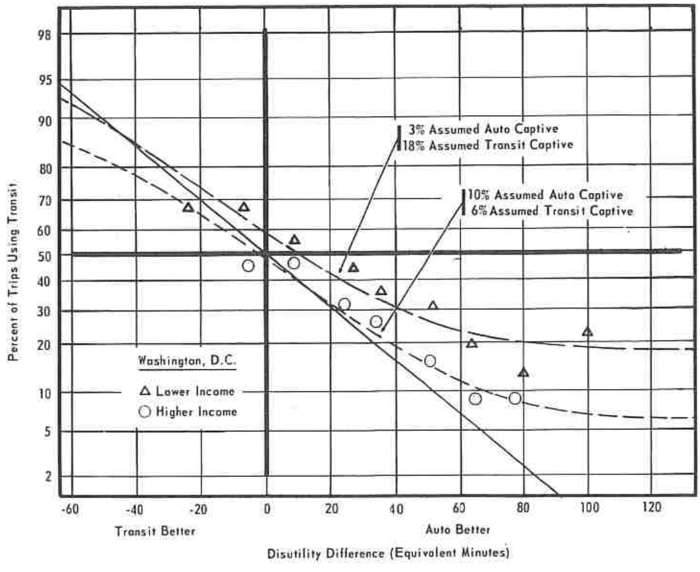


Figure 9. Washington modal-choice relationships.

The analysis of the utilitarian theory of mode choice has been carried one step further by adjusting both the Twin Cities and the Washington data to remove the assumed captive trips from consideration. This was done using the following computation.

$$\text{Free choice percent transit} = \frac{100\ [(\text{overall percent transit}) - (\text{percent transit captives})]}{100 - [(\text{percent transit captives}) + (\text{percent auto captives})]}$$

The results of analyzing the free choice trips alone are shown in Figure 10 as plotted on normal probability paper. The vertical axis represents the percentage of noncaptive trip-makers who choose transit. All data points from both cities and all income groups are included. Occasional negative values result from the observed transit usage being below the estimated transit captivity percentage.

Use of the normal probability coordinate system on the vertical axis amplifies the dispersion of data points at the larger positive and negative disutility difference values. To assist in evaluating the results, corresponding curves at ±5 percentage points have been provided in addition to the principal hand-fitted predictive curve.

Figure 10 was prepared by plotting the more extensive Twin Cities data first, thus establishing the curve. The Washington data were then added on a trial basis prior to setting the final captivity percentages. These percentages were hand adjusted until the best fit was obtained. It was in this manner that the Washington captivity percentages were established.

The exercise serves to show how the theory might be applied to limited data and to a model calibrated to fit an expression predetermined elsewhere in a more comprehensive study. It is not known at this point, of course, if more information on the 1955 Washington travel patterns would substantiate the captivity percentages chosen. The higher transit captivity and lower auto captivity postulated for Washington is consistent, at least, with the comparative urban characteristics of the two cities.

If it were known with certainty that the Washington captivity probability percentages are correct, then the evidence would clearly indicate that not only is the normal probability density function applicable to both cities but that the variance is the same. This would, in turn, mean that the noncaptive populations of the two cities make the choice between transit and auto according to a common, and therefore presumably universal,

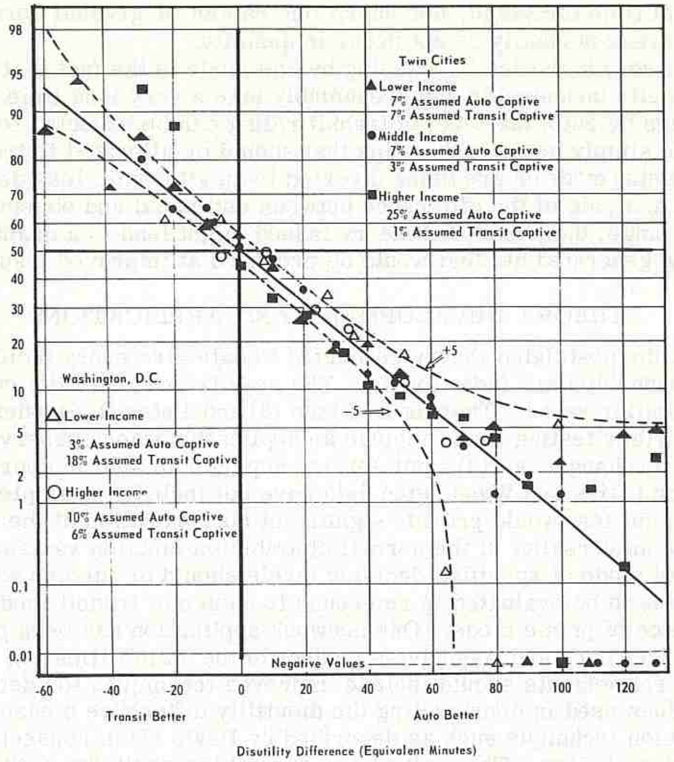


Figure 10. Twin Cities and Washington free choice of mode.

relationship. Final judgment, however, should be withheld until comprehensive data are available from another urban area. It should be noted that the response to disutility difference clearly differs in variance at different levels of decision. This can be seen by comparing the submodal-split results with the prime modal-split examples.

Use of the single curve for free choice of mode at all economic levels does not imply that income has no influence. Income is a determinant of the relative dollar value of time used in constructing the disutility values. The extent of the resultant effect of income on modal choice depends in part on the characteristics of the transit system. Given a situation where transit is consistently cheaper but slower than auto, rising incomes will cause a marked decline even in free choice patronage as forecast by the curve. This, of course, parallels the general experience of surface transit operations over the past two decades of increasing affluence.

EXCESSIVE TRIP LENGTH EFFECTS

As was alluded to previously, the captivity effect is joined by a second significant deviation from the normal distribution. This additional effect takes the form of lack in observed as compared to estimated transit riding at high disutility difference values unfavorable to transit. In the Twin Cities case, very little transit riding is observed at difference in disutility values over 130 equivalent minutes unfavorable to transit regardless of the postulated constant probabilities of transit captivity. Comparable effects are common in the results of other modal-split modeling work.

A logical explanation would appear to be that the deficiency in observed transit trips is a reflection of the overall metropolitan area trip generation and distribution characteristics. Trip-length analyses covering both auto and transit travel indicate that

there is a travel time threshold, not sharp but rather of gradual form, above which trips in urban areas normally do not occur in quantity.

Implicit in a very high disutility saving by one mode is the fact that the less favored mode must be quite undesirable and presumably take a very long time. Thus, at high disutility savings by auto, the level of transit riding can be expected to drop below the estimated value simply because the trips that should be allocated to transit are in reality either not being made or are being diverted to an alternate, less desirable, destination. Through analysis of the difference between estimated and observed transit usage in the critical range, the postulated theory indeed might lead to a quantitative measure of trips not now generated but that would be produced at improved transit service levels.

THEORY DEVELOPMENT AND APPLICATIONS

The tests of the postulated theory conducted to date give every indication that the suggested relationships are indeed valid. The results are generally compatible with work along a similar vein by Thomas E. Lisco (8) and Peter R. Stopher (9).

Desirable further testing would include an application where observations could be made of trip interchanges with transit service superior to auto in door-to-door travel time. The Twin Cities and Washington data have not included examples of this condition. A successful test would provide significant corroboration of the postulated theory.

The possible universality of the normal distribution function variance in application to free choice of mode at specified decision levels should be further explored. The theory also needs to be evaluated in reference to choice of transit mode of access and to nonwork choice of prime mode. One nonwork application has been performed as part of the Alan M. Voorhees and Associates project in the Twin Cities (4).

Operational refinements should include improved techniques for determining the equivalence values used in constructing the disutility difference measures, perhaps using a regression technique such as described by Bevis (3) in connection with road-user cost function studies. The method for describing captivity probabilities also needs attention. It should be feasible to estimate captivity probabilities on the basis of trip-maker and metropolitan area characteristics. It should also be possible to predict deficiencies in the number of long transit trips by reference to the implications of the trip distribution model being used.

The postulated theory has immediate application for estimating modal split and route choice within the standard travel forecasting sequence. It should also be applicable to simultaneous multimode assignment upon development of an appropriate multipath assignment technique.

A further interesting possibility is that a comparable theoretical description of travel activity may be possible on a broader scale. If the modal choice and route choice facets of urban travel can be described by utilitarian behavior, disutility measures, and probability theory, perhaps trip generation and distribution can be handled similarly. This could be a fruitful area for detailed research and evaluation.

CONCLUSION

It is concluded that choice of travel mode can be treated as an economic response to the transportation system characteristics. As such, it can be predicted using quantitative disutility comparisons and probability mathematics.

The proposed structure of the descriptive equation set forth gives a logically possible explanation of modal choice and appears to produce good results even with the thin data inherent at the trip interchange level of analysis.

Although further testing and refinement would be highly desirable, the results and the credible framework of the postulated utilitarian theory would indicate that it is readily applicable at the present time to forecasting modal-split and submodal route choice. It is further thought that the basic concepts should be investigated for usefulness in describing other aspects of urban travel.

ACKNOWLEDGMENTS

The formulation of the material in this paper and the preparation of the text were significantly aided by several interested persons who, at various stages, contributed valuable data, suggestions, encouragement, and criticism. Particular mention should be made of Gordon Schultz, Gordon Shunk, Thomas Deen, Frank Navin, and Howard W. Bevis. The author wishes to thank them all.

REFERENCES

1. Quarmby, D. A. Choice of Travel Mode for the Journey to Work: Some Findings. Jour. Transport Economics and Policy, Vol. 1, No. 3, Sept. 1967.
2. Pratt, Richard H., and Deen, Thomas B. Estimation of Sub-Modal Split Within the Transit Mode. Highway Research Record 205, 1967, pp. 20-30.
3. Bevis, Howard W. Estimating a Road User Cost Function From Diversion Curve Data. Highway Research Record 100, 1965, pp. 47-54.
4. Alan M. Voorhees and Associates. Work performed for the Twin Cities Area Metropolitan Transit Commission, 1969.
5. Shunk, Gordon A., and Bouchard, Richard J. An Application of Marginal Utility to Travel Mode Choice. Paper presented at the 49th Annual Meeting and published in this RECORD.
6. Ginn, J. Royce. Unpublished work at Alan M. Voorhees and Associates, 1966.
7. Traffic Research Corporation. A Model for Estimating Travel Mode Usage in Washington, D. C., Volume 2, Listings of 1955 Summarized Data. Prepared for the National Capital Transportation Agency, July 1962.
8. Lisco, Thomas E. The Value of Commuter's Travel Time: A Study in Urban Transportation. Univ. of Chicago, unpublished PhD dissertation, 1969.
9. Stopher, Peter R. A Probability Model of Travel Mode Choice for the Work Journey. Paper presented at the 48th Annual Meeting of the Highway Research Board, January 1969.

River Crossing Travel Choice: The Hudson River Experience

EUGENE J. LESSIEU, The Port of New York Authority; and
JEFFREY M. ZUPAN, Regional Plan Association

A set of models has been developed to allocate peak-period and off-peak-period trips for each of three modes—auto, bus, and rail—to facilities crossing the Hudson River. The set of allocation models is one of a series of models to be used to forecast trans-Hudson travel for alternative transportation systems within the framework of the New York region's changing demography and economy. Previous techniques were reviewed to determine the best one for developing the allocation or assignment process. Unlike most previous techniques, the method selected incorporated a number of determinants of route choice. Multiple regression analysis fed by a massive data bank and a large battery of programs was used. Times, costs, and number of transfers were compared on an origin-destination basis for each crossing facility within a mode; and their relative transportation parameters were ascertained to describe variations in facility usage. The results showed the great influence of time savings on the auto user, suggesting the general validity of the often-used all-or-nothing minimum-time-path approach to assigning auto traffic. The allocation models for the other modes suggest a lesser but still high value of time, with the differential number of transfers being an important determinant of the rail crossing choice. Dummy variables to test user biases toward particular facilities were tried but with no usable results. Families of curves of the models were prepared that greatly aided the analysis and understanding of the models. The allocation models were run for the base year 1964 to compare the results with the actual trip volumes. Results were generally good, but "fine tuning" was necessary for the auto mode. The models developed were deemed usable for forecasting purposes with full knowledge of the limitations of such empirically derived relationships. A continuing research effort with new data and improved techniques is being planned.

•THE PORT of New York Authority is engaged in the development of a series of traffic-demand forecasting models to aid in the planning of transportation facilities related to crossing the Hudson River. The goal of the model development is to efficiently forecast the trip demand by mode and facility of alternative transportation plans and policies within the framework of the changing demography and economy of the New York metropolitan area. Toward that end, a system of models was developed that is described schematically by the flow chart in Figure 1. There are three basic submodels in the in the system: a trip interchange model that forecasts the total number of trans-Hudson trips made between zonal pairs; a modal-split model that apportions these trips among the three major travel modes—automobile, bus, and railroad; and an allocation model that apportions the modal trips to each facility within the mode used to cross the river. It is this last model, the allocation or assignment model, that is the subject of this study.

Paper sponsored by Committee on Origin and Destination and presented at the 49th Annual Meeting.

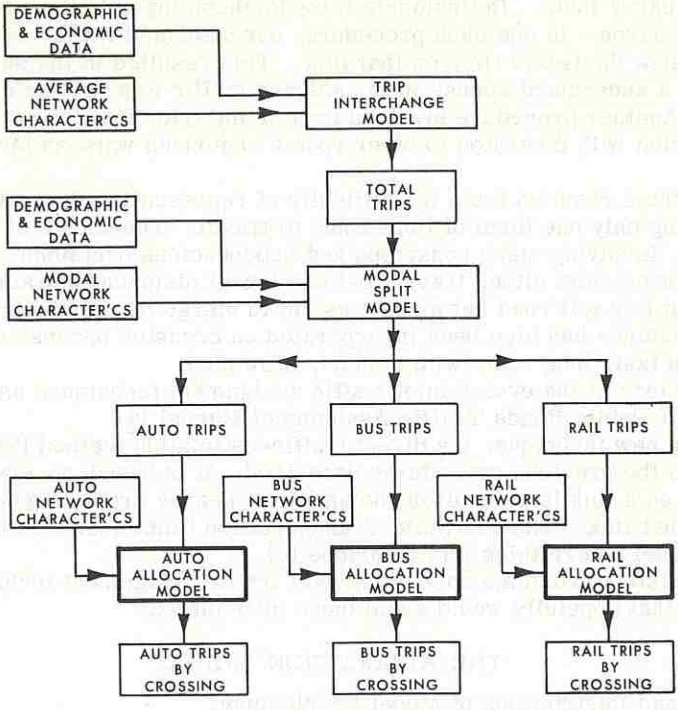


Figure 1. The model system—schematic flow chart.

DEVELOPMENT OF ASSIGNMENT TECHNIQUES

The techniques of forecasting the traffic assigned to a transportation network have evolved considerably in the last 15 years. In the early years of highway planning, "desire lines" were drawn between expected travel interchange points in proportion to the thickness of the volumes, and highway locations were then sketched in.

Later, traffic was assigned to expected routes of travel empirically derived with the aid of diversion curves. The relative time and/or distance and/or cost savings were calculated for the added facility, and the percent of automobile travelers that would switch over to the new facility was calculated for each origin-destination (O-D) zonal pair (1).

The increased availability and use of high-speed digital computers and the presentation of the Moore algorithm (2) in 1957 helped to improve the process of traffic assignment. It became possible to trace the route of least time, distance, or cost through a transportation network described in a computer and to assign each O-D trip volume to links in the network describing that route. Once accumulated, the trips on each link represented an estimate of the traffic that would be assigned.

There remained a number of serious shortcomings, however, with this assignment method. First, traffic was assigned on an all-or-nothing basis. All traffic was assigned to the minimum route over all other possible routes, no matter how small the margin. This was unrealistic because motorists will choose the next best route in significant numbers if the margin is small. This problem was overcome somewhat by the use of small traffic zones, thereby smoothing the lumpiness of the assignment. This solution, of course, added to the number of O-D pairs.

A second problem occurred because the computer was unaware of the phenomenon of traffic congestion. Traffic was assigned to links in the network that exceeded the capacity

of the links to carry them. Refinements were forthcoming with the advent of capacity restraint procedures. In one such procedure, overassigned traffic on a link caused the computer to raise the travel time on that link. This resulted in the selection of other routes and in a subsequent reduction in assigned traffic to a volume closer to the link's capacity (3). Another procedure involved loading the trips into the networks in an incremental fashion with diversion to other routes occurring when an increment produced overloading.

A third problem resulted from the difficulty of representing the travelers' preference for a route using only one form of impedance to travel. The choice of route is indeed a complex one, involving many conscious and subconscious decisions. Although travel time has been used most often, travel cost and travel distance have also been considered. The assignment to a toll road has made it useful to merge time and cost by using some equivalent. Distance has also been incorporated on occasion because of the problems of comparing a fast, long route with a short, slow one.

A sketch history of the evolution of traffic assignment techniques as of 1964 is found in the Bureau of Public Roads Traffic Assignment Manual (4).

A promising new technique, the direct traffic-estimation method (5), is a complete departure from the previous procedures described. It is based on the concept that traffic volume on a link is a result of the ability of nearby areas to generate trips plus the access to that link. The Tri-State Transportation Commission is currently calibrating this model and refining this technique (6).

Given the existing problems associated with traffic assignment techniques, a method was attempted that hopefully would avoid these difficulties.

THE ALLOCATION MODEL

Requirements and Restrictions of Model Development

The Port of New York Authority's approach to the problem of assignment or allocation (the latter term will be used hereafter) is governed by the unique nature of both its responsibilities and its data base. The Port Authority's concern is with the Hudson River crossings and the facilities that directly affect them. The allocation process to be devised must focus on these facilities and must be applicable to the three primary modes of trans-Hudson travel—auto, bus, and rail. The Port Authority has collected a great deal of O-D information on these trans-Hudson crossings and has coded them in some areas to what might be considered a gross zone base.

The more traditional approaches to assignment require the construction of extensive networks, a process which is rather wasteful if only a few links on the network (i.e., the Hudson River crossings) are of concern. In addition, with the data base, the traditional all-or-nothing approach would result in a great deal of lumpiness in the assignment, especially in the bus and rail systems where relatively few zones contribute a large portion of the trips.

It was also thought that the choice of a route across the river was based on more than just one variable, particularly in the bus and rail modes. For the bus mode, in a portion of the region west of the river, the trip-maker has a choice of two sets of bus lines, each to a different bus terminal on the east side of the river. These alternatives present a number of choices of trade-offs for the trip-maker involving travel time differences, travel cost differences, and a differential number of transfers. There can be four basic alternatives for traveling by rail from the zones on the west side of the river. They also involve many trade-offs for the trip-maker. Five different automobile

TABLE 1

HUDSON RIVER FACILITIES BY MODE

Mode	Facility
Auto	George Washington Bridge (GWB) Lincoln Tunnel (LT) Holland Tunnel (HT) Staten Island Bridges (SIB) (Bayonne Bridge, Goethals Bridge, Outerbridge Crossing) Tappan Zee Bridge (TZB)
Bus	George Washington Bridge Bus Station (GWBBS) Port Authority Bus Terminal (PABT)
Rail	Pennsylvania Railroad (PRR) Hudson Terminal (HT) Port Authority Trans-Hudson (PATH) Uptown (PUP) Central Railroad of New Jersey (CNJ)

crossings are also available. The 11 choices for the three modes are given in Table 1 and are shown in Figure 2.

Considering the difficulty of constructing huge networks in this situation, the lumpiness of all-or-nothing assignments, and the difficulty of using one variable to describe the travel impedance, it was decided to develop an allocation model designed to overcome these difficulties.

The Model Concept

The allocation technique employed is based on the concept that each crossing facility within a mode of travel competes with all others for the trips made within that mode between each O-D pair. Although it is true that there is competition between modes as well as between facilities within a mode, a considerable amount of literature indicates that different factors govern modal choice. These factors might not be handled easily in an allocation method that does not specifically identify the mode. The technique considered for allocation within modes does not necessarily identify the facility per se in its concept.

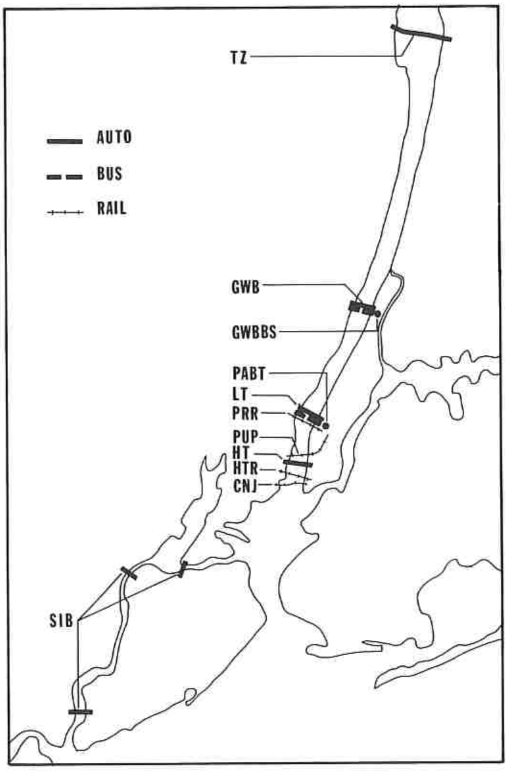


Figure 2. Hudson River crossings map.

The allocation model is based on a rating system first introduced by Cherniack (7). The concept assumes that the traveler compares the travel time, travel cost, and (in the case of bus and rail) the number of transfers for the various available facilities. In evaluating the alternatives, the traveler perceives the fastest facility and compares that time to the times of the other facilities; he perceives the least expensive facility and compares that cost to the costs of the other facilities; he perceives the most convenient alternative and compares it to the others; or, more realistically, he perceives some combination of all factors. He then rates the alternate facilities and gives the highest rating to the one that he decides has the best combination of time, cost, and convenience and a lesser rating to those he believes lack these advantages. Conversely, if the use of each facility is based on the cumulative rating of all users, then each facility could be given a rating based on its traffic volume compared with the traffic volume of all other competing facilities. The facility with the highest volume gets the highest rating; and others, comparatively lower ratings.

Using multiple regression techniques, the relationship between these three factors and the comparative usage of the facilities was explored for each mode. The rating of facility 1 can be expressed as follows:

$$R_1 = \frac{T_1}{T_H} = f\left(t_1 - t_s,\ c_1 - c_c,\ F_1 - F_f\right)$$

where

T_1 = trips via facility 1,
T_H = trips via facility most heavily used,
t_1 = door-to-door travel time via facility 1,
t_s = door-to-door travel time via the fastest facility,
c_1 = travel cost via facility 1,

c_c = travel cost via the least expensive facility,
F_i = number of transfers via facility 1, and
F_f = number of transfers via the facility with the fewest transfers.

The R value or rating will equal 1.0 if the facility in question is the most heavily used and will be less than 1.0 for all lesser-used facilities. Also, the differences will equal zero if the facility in question is the best for the particular transportation variable. The ratings and the differences (Δt, Δc, and ΔF for times, costs, and transfer differences respectively) are calculated for each facility within each O-D pair for each mode. Thus, for the automobile allocation model with five available crossings, each O-D pair can theoretically contribute five data points. In this study, each O-D pair contributed fewer points because only those facilities that were within 20 minutes of the fastest were deemed worth considering. Needless to say, few if any trips were found in that excluded category.

When using the model to forecast facility usage, it is not necessary to find the most heavily used facility. The rating for each facility, being the dependent variable, is determined by the time, cost, and transfer differences. The share of the total traffic for each facility is the ratio of its rating to the sum of all the ratings.

Input Data Development

A few words are in order concerning the problems of data collection and handling. The Port Authority analysis zones (Fig. 3) were used. On the west side of the Hudson River, 92 zones were considered; on the east side, 69 zones. Included, then, were 6,348 O-D pairs. The entire model system was designed to consider only average weekday travel. The calibration process was based on 1964 data. The peak period (7 a.m. to 10 a.m.) was considered separately from the off-peak period. Travel times and travel costs had to be found for each of the O-D pairs for each time period for each of the 11 crossing facilities considered. In addition, six facilities required transfer values. To be added to this were the trip volumes for each cell, for each time period, and for each facility. The items of data totaled 494, 544, and therefore a high-speed digital computer was employed with a data bank and a battery of supporting programs having great flexibility.

The determination of the proper values to be placed in the data bank merits some attention. Auto trip data were taken from the continuous-sample O-D surveys taken at the Port Authority facilities and at the Tappan Zee Bridge. Bus trip data were based on O-D surveys taken at the two bus terminals. Rail trip data, including the PATH system, were synthesized from a PATH O-D survey, from O-D surveys of those rail lines involved in the Aldene Plan (Central Railroad of New Jersey; Pennsylvania Railroad, Shore Branch), from various conductor counts, and from the Manhattan Journey-to-Work Surveys taken in 1961-1962.

For auto times and costs, it was necessary to build peak and off-peak link-and-node networks. Travel time for each facility was calculated along the minimum-time path with all of the other trans-Hudson facilities removed from the system. Costs were found by skimming over those paths and were based on over-the-road costs of 2.8 cents per passenger-mile plus tolls and average parking costs.

The bus and rail time, cost, and transfer matrices were developed by adding rows and columns for what might be called a common-point network. Travel times were determined from each zone west of the Hudson to a Manhattan terminal (Penn Station, for example). Travel times then were determined from that terminal to each zone east of the Hudson. The same was done for costs and transfers. This depicted quite naturally how a bus or rail trip is made; and it was necessary only to add the rows and columns to determine the full i to j matrix of all time, cost, and transfer data.

The theory that travelers would show a preference toward a particular route, even if it was not superior according to our measures, was also set up for testing. By using a dummy variable for each facility, it was possible to determine if there was a significant bias toward a particular facility. It was theorized, for example, that bridges were preferred to tunnels, irrespective of small time and cost differences.

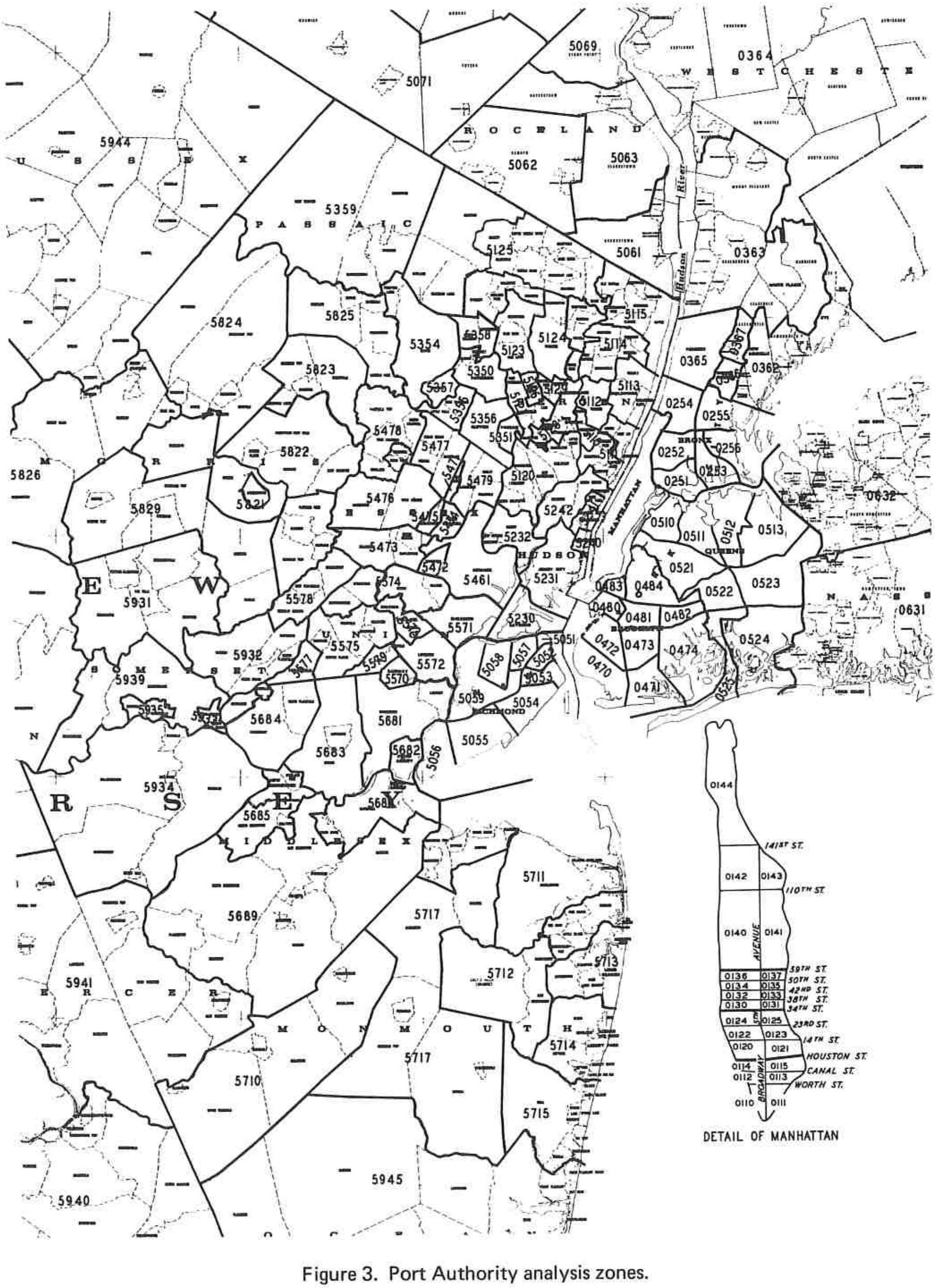


Figure 3. Port Authority analysis zones.

Model Development and Results

Many multiple-regression trials were run for each of the three travel modes. Data points were weighted in proportion to the modal trips in the O-D pairs, so that the less statistically reliable low-volume points were not heavily influential. The trials involved (a) testing linear and curvilinear forms, (b) the inclusion and the exclusion of the facility dummy variables, and (c) a further stratification of the data into trips oriented toward the central business district (CBD) and non-CBD-oriented trips. The resultant regression equations were studied for reasonable size and correct sign of coefficients. The final tests involved the application of the most promising equations to the total modal base-year trips and a subsequent analysis of the resulting differences from the sample O-D trip pattern to determine whether they reproduced the base year reasonably.

The chief findings of these regression trials were as follows:

1. The best form of the equations for the multiple correlation coefficient was $R = \exp\left[b1\Delta t + b2\Delta c + b3\Delta F + K\right]$;
2. Time differences were clearly the most significant determinant of route choice, particularly for the auto mode;
3. Cost differences were consistently the second most significant determinant;
4. Transfer differences were only significant for the rail model;
5. The CBD equations were significantly different from the non-CBD equations, with the exception of peak-period auto mode;
6. The auto allocation equations reproduced the assignment to the auto crossings fairly well but required small amounts of "fine tuning";
7. The bus allocation equations for the peak period did well in total assignment but were the result of large errors in isolated zones canceling one another out; those for off-peak period assigned very well; and
8. The rail equations assigned fairly well.

The vital statistics for the equations finally selected are given in Table 2. They are all of the exponential form just described. One equation was satisfactory for both the peak CBD and the peak non-CBD. When tried separately, the results were almost identical. Otherwise, the stratifications tried for each mode were significantly different from one another.

The facility dummy variables either indicated no biases that fit the possible theories previously set down or else did not improve the accuracy of the forecasting process. The authors had decided beforehand that for the auto mode there might exist a built-in preference for bridges rather than for tunnels. The results, however, were a crazy-quilt pattern of relatively insignificant coefficients of the dummy variables that neither confirmed this theory nor suggested a new one. For the rail mode, it was theorized that biases would favor the commuter railroads over the PATH facilities. Again, no

TABLE 2

ALLOCATION MODEL DATA

Mode	Time	Orientation	Weighted N (degrees of freedom)	Coefficients			K (constant)	R
				Δ Time (min)	Δ Cost (cents)	Δ Transfers		
Auto	Peak	CBD + non-CBD	9,267	-0.536	-0.073	—	-1.915	0.660
	Off-peak	CBD	5,367	-0.695	-0.038	—	-2.052	0.712
	Off-peak	Non-CBD	7,080	-0.624	-0.050	—	-1.277	0.688
Bus	Peak	CBD	4,817	-0.249	-0.063	—	+0.085	0.525
	Peak	Non-CBD	615	-0.386	-0.227	—	-0.781	0.481
	Off-peak	CBD	1,978	-0.324	-0.091	—	-0.275	0.450
	Off-peak	Non-CBD	1,099	-0.427	-0.160	—	-0.534	0.539
Rail	Peak	CBD	8,353	-0.360	-0.081	-1.399	-0.030	0.659
	Peak	Non-CBD	767	-0.557	—	—	-2.821	0.521
	Off-peak	CBD	2,843	-0.306	-0.077	-0.470	-0.642	0.469
	Off-peak	Non-CBD	1,510	-0.438	-0.162	—	-1.297	0.477

TABLE 3

VARIABLE EQUIVALENCES

Mode	Time	Orientation	Value of Time (cents/min)	Value of Transfer (cents)	Time Value of Transfer (min)
Auto	Peak	CBD + non-CBD	7.3	—	—
	Off-peak	CBD	18.3	—	—
	Off-peak	Non-CBD	12.5	—	—
Bus	Peak	CBD	4.0	—	—
	Peak	Non-CBD	1.5	—	—
	Off-peak	CBD	3.6	—	—
	Off-peak	Non-CBD	2.7	—	—
Rail	Peak	CBD	4.4	19.7	3.9
	Peak	Non-CBD	No equivalences, only time considered		
	Off-peak	CBD	4.0	—	—
	Off-peak	Non-CBD	2.7	6.1	1.5

clear pattern emerged. Only for the bus mode, where it was theorized that riders would prefer the PABT over the GWBBS, did some semblance of expected preferences hold. For the peak CBD bus model, a dummy variable showing such a preference entered the equation. When this equation, however, was used in an attempt to reproduce the base-year trips, it did not perform as well as the equation without the dummy variable.

Interpretation of Model Results

The variable equivalents given in Table 3 must be interpreted with great caution. Because of the nature of the exponential decay form of the models, these equivalencies are only applicable at the lower ranges of the independent variables. They do not apply at the higher ranges where the curves approach the ratings asymptotically. Table 3 does give some interesting information, however. It indicates that the automobile user places a greater value on time than does the public transportation user. This appears to be logical because his choice of the auto mode in the first place generally reflects his interest in time savings and his lack of concern for high costs. The lower value of time exhibited by the peak auto users as compared to the off-peak users also seems reasonable. This might reflect the user's perception of the accumulation of costs over five round trips each week that are common to the peak auto user. Presumably, saving a few cents each day is important enough to be acted upon. The off-peak auto user is more likely to be the occasional trans-Hudson traveler. In such cases, the most direct and fastest route is apparently considered first, and the occasional extra toll is not paid often enough to be weighed heavily in route selection.

The significance of transfers for the rail mode is worthy of note. For rail trips made to the CBD in the peak period, the equation states that when various routes present roughly equivalent choices, the elimination of one transfer will attract as much traffic as the decrease in time of about 4 minutes or the lowering of the fare by about 20 cents.

The series of models are shown as graphs in Figures 4, 5, and 6. The series of curves on the left column of graphs show rating versus Δc for a family of curves of Δt. The other graphs show the same data in the form of rating versus Δt for a family of curves of Δc. The rating scales for the curves were normalized to make the rating equal to 1.0 where Δt, Δc, and ΔF all equaled zero in order to simplify description of the meaning and use of the curves. The meaning of these curves can best be described by examples.

Consider a case where three auto facilities are available for a peak-period trip from i to j. Facility 1 requires a travel time of 45 minutes at a cost of 1 dollar; facility 2, 47 minutes and 90 cents; facility 3, 50 minutes and 80 cents. Time and cost differences would be calculated from the least time and the lowest cost, and the rating would be read from the graph as shown in Figure 7. The percentage of total traffic from i to j that

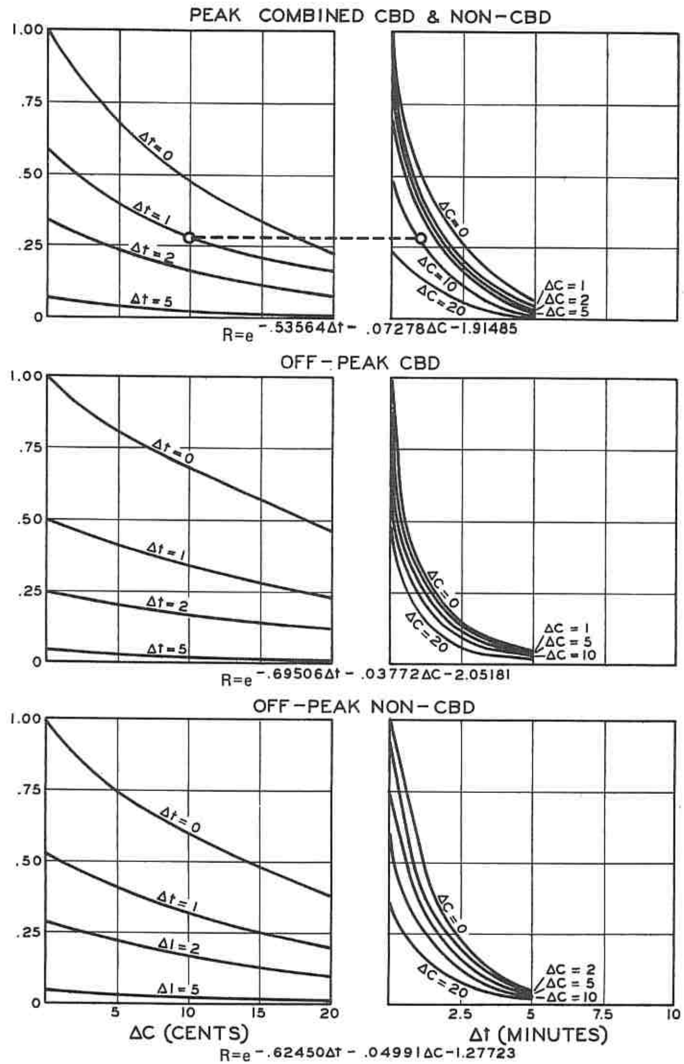


Figure 4. Auto allocation models.

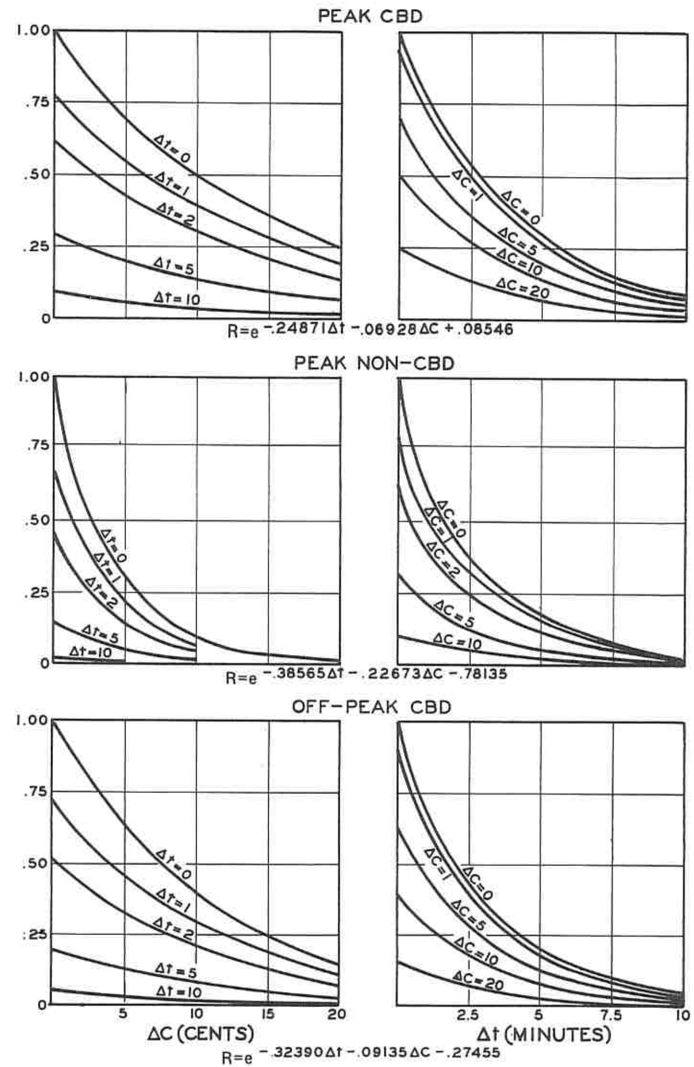


Figure 5. Bus allocation models.

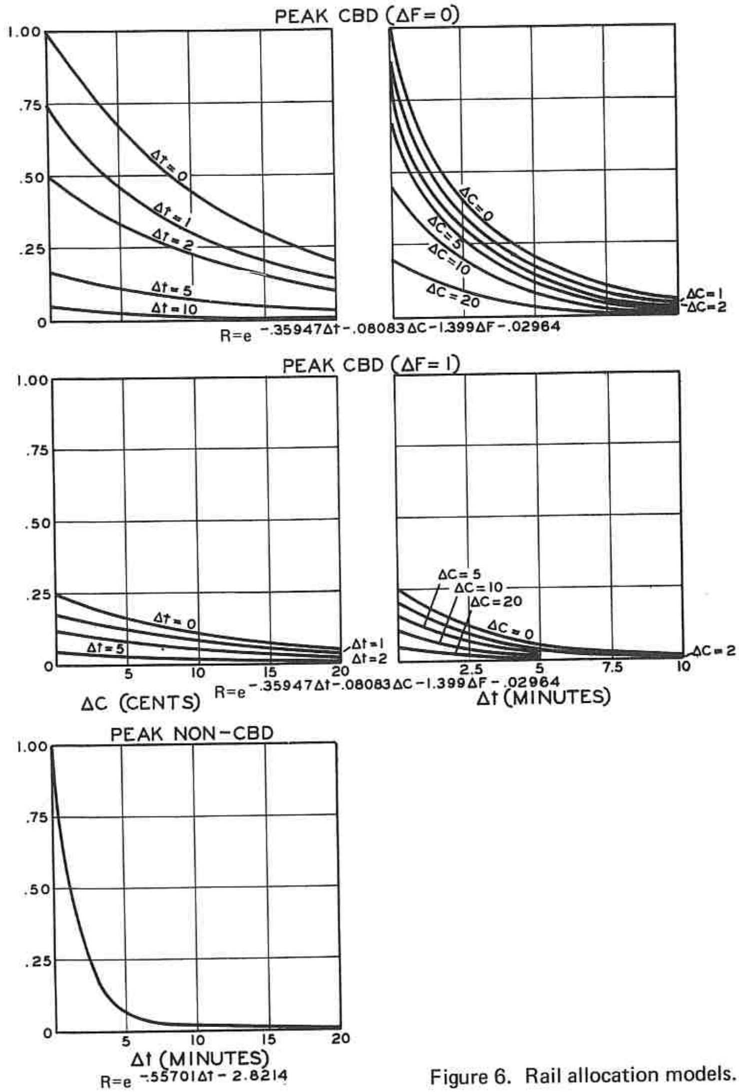


Figure 6. Rail allocation models.

each facility would be assigned would be calculated by dividing its rating by the sum of all ratings as given in Table 4. Note that the fastest but most expensive facility received the largest share of the traffic. Should travel via facility 1 then be slowed by only 2 minutes, the redistribution of traffic as in the second group of data would result. In this case, facility 2 captures the largest share of the traffic. This example indicates that the auto user will prefer the fastest facility even if it is more expensive. Only when times are nearly identical will cost become the determining factor.

Another example can be shown using the peak CBD rail curves (Fig. 8). Assuming three competing facilities with travel times of 60 minutes, travel costs of 60, 80, and 70 cents respectively, and with one additional transfer required for travel via the third facility, we obtain the distribution given in Table 5. If travel via facility 3 were made direct, without a transfer, the second group of results in Table 5 would be obtained. The removal of that transfer has enabled facility 3 to more than triple its share of the traffic. It also can be seen that facility 1 retains the majority of the traffic on the basis of its lower cost alone.

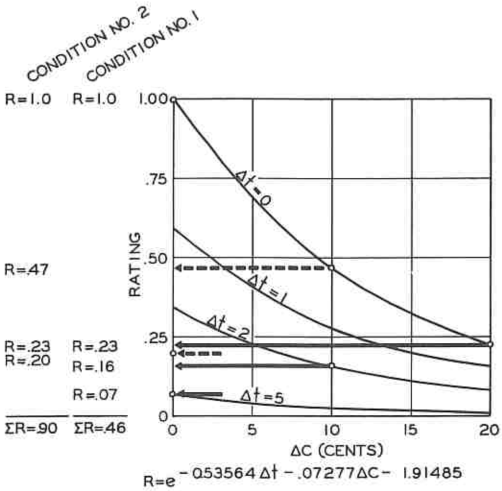


Figure 7. Peak auto model (combined CBD and non-CBD) example of use.

TABLE 4

CURVE APPLICATION: EXAMPLE 1

Facility	t	c	Δt	Δc	R	Share (percent) (R/ΣR)
1	45	\$1.00	0	20	0.23	50
2	47	0.90	2	10	0.16	35
3	50	0.80	5	0	0.07	15
					ΣR 0.46	
1	47	1.00	0	20	0.23	26
2	47	0.90	0	10	0.47	52
3	50	0.80	3	0	0.20	22
					ΣR 0.90	

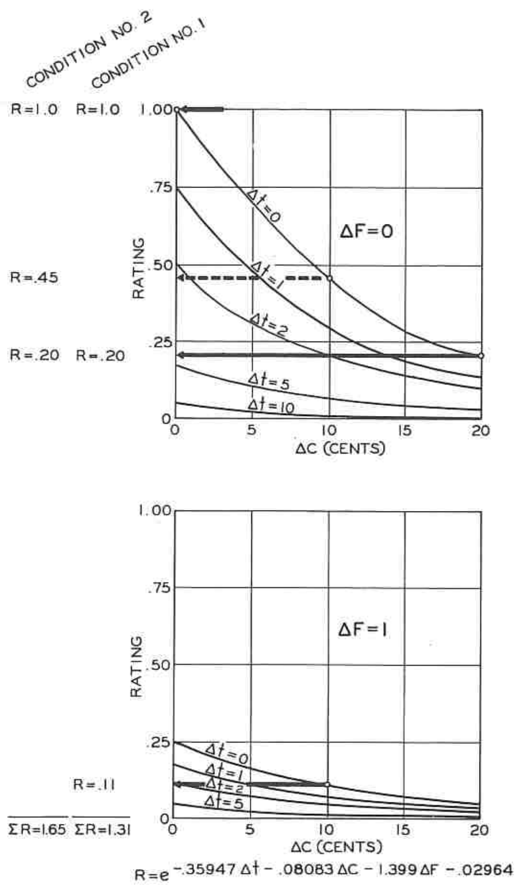


Figure 8. Peak rail model (CBD) example of use.

Trial Runs of the Models

Before it is possible to accept the model as a forecasting tool, it is necessary to see how well it predicts the base-year trips. This was done for a series of models; the end result was the acceptance of the models described but with some adjustments. In analyzing the results of the predictions, the volumes assigned to each facility were compared

TABLE 5

CURVE APPLICATION: EXAMPLE 2

Facility	t	c	F	Δt	Δc	ΔF	R	Share (percent) (R/ΣR)
1	60	60	1	0	0	0	1.00	77
2	60	80	1	0	20	0	0.20	15
3	60	70	2	0	10	1	0.11	8
							ΣR 1.31	
1	60	60	1	0	0	0	1.00	61
2	60	80	1	0	20	0	0.20	12
3	60	70	1	0	10	0	0.45	27
							ΣR 1.65	

TABLE 6

COMPARISON OF ASSIGNMENTS: AUTO

Method	Number of Trips					
	GWB	LT	HT	SIB	TZB	Total
Peak:						
Actual	23,560	11,159	3,214	2,770	5,205	45,909
Minimum time	24,868	11,106	2,679	2,424	4,834	45,909
Model	24,993	10,612	3,179	2,535	4,590	45,909
Model "tuned"	23,675	11,677	3,269	2,540	4,749	45,909
Off-peak:						
Actual	62,301	33,045	25,013	14,808	8,916	144,083
Minimum time	76,626	28,270	19,195	13,809	6,595	144,493
Model	72,892	29,087	22,124	14,175	6,259	144,489
Model "tuned"	63,883	34,766	24,067	14,235	7,259	144,214

to the actual volumes. However, this was not sufficient. It was also necessary to compare the results at a finer grain to determine whether the county-to-county or even zone-to-zone volumes compared well. It is at this level that problems were uncovered. Tables 6 to 8 give the total volume comparisons. In each case, the actual trips are shown on the first line and the trips assigned via the minimum-time-path method are shown on the second line. The latter was tried in order to examine the results that would be obtained by the more standard all-or-nothing method. Subsequent lines show the results of the key runs of the models.

For the auto assignments (Table 6), the model produced assigned volumes similar to the minimum-time-path assignment for the peak and performed somewhat better in the off-peak. This indicated what we had already come to know. The auto user is so heavily influenced by time that the minimum-time path is not all that bad. Nevertheless, having cost in the model does play some part in the assignment and makes the model available for testing cost or price changes. "Fine tuning" of the auto models involved an addition of an arbitrary time delay at one facility that was consistently overpredicted. Although this might have been considered a network correction, we could, in all honesty, find no justification from the observed field data to make this change.

The bus assignments are given in Table 7. As described earlier, the peak bus model had been tried with a dummy variable to explain the preference for the PABT. The data in Table 7 show that both

TABLE 7

COMPARISON OF ASSIGNMENTS: BUS

Method	Number of Trips		
	GWBBS	PABT	Total
Peak:			
Actual	11,268	47,577	58,845
Minimum time	8,972	49,873	58,845
Model with dummy	9,776	49,068	58,845
Model without dummy	10,651	48,194	58,845
Off-peak:			
Actual	10,706	32,366	43,072
Minimum time	9,222	33,850	43,072
Model	10,548	32,510	43,058

TABLE 8

COMPARISON OF ASSIGNMENTS: RAIL

Method	Number of Trips				
	PRR	HT	PUP	CNJ	Total
Peak:					
Actual	7,593	26,060	13,153	7,878	54,684
Minimum time	8,050	24,864	12,548	9,196	54,658
Model	7,442	25,962	12,532	8,722	54,658
Off-peak:					
Actual	2,648	9,854	5,998	390	18,890
Minimum time	1,341	10,136	6,616	798	18,890
Model	2,842	9,669	5,218	988	18,729

the use of this dummy variable and the use of the minimum-time path did not yield as close an agreement with the actual volumes as did the model selected. The data are deceiving, however. The close agreement of the model in total masks some large zones that fit very poorly but cancel one another out. It was concluded that the differential frequency of service is probably the factor missing from the allocation model. The off-peak bus model results fit very well in total as well as at the zonal level.

The rail assignments are given in Table 8. The minimum-time path assignment, which looks good in total, particularly for the peak, was very poor when individual zones were observed. This is no surprise because the model has shown that cost and transfers can play a significant role. The model assigned extremely well with only small difficulties related to the CNJ. These were in large zones or in zones where parallel competing services existed. In both instances, it was impossible to accurately describe the differences in service with the variables used. Because the CNJ system has been drastically revamped since the base year, the inaccuracy was of minor concern.

MODEL IMPLICATIONS

The modeling effort just described has provided us with a forecasting tool and with some knowledge concerning the relation of transportation network variables and trip route choice. Recognizing the imperfectness of fit in these relationships, however, can also be considered as knowledge gained.

The fact that auto assignment from the model differed only slightly from an all-or-nothing minimum-time assignment indicates that minimum time might be a reasonable method of allocating auto trips. It also shows that the model developed is highly reactive to time changes and is much less so to cost changes. Although this corroborates the results of other investigations on this subject, the exact value of the cost coefficient and the relation of the cost and time variables should be treated very carefully.

The bus allocation models showed that both time and cost differences proved significant, with the latter having far greater effect than they did in the auto model. Some index of convenience, however, such as frequency of service might have added to the quality of these models. Unfortunately, the data base that governed the zone description made it impractical to accurately describe a frequency variable.

The rail allocation models suggest a relation between time and cost similar to that of the bus allocation models. They further suggest a relatively strong reaction to the number of transfers. Because the transportation system from which the models were derived frequently requires at least one transfer (to the New York City subway system) and because this transfer involves additional cost, it is suggested that cost and transfers might be closely associated. Nevertheless, based on the rail system in this area, the model suggests that a rider might be more willing to pay an additional cost than to add time to his trip to avoid a transfer; and this appears to be reasonable.

The reader should remember that the models described herein are route-choice models, not modal-choice models.

In using any of the relationships, the modeler must be aware of the assumption implicit in all such efforts. There is no guarantee that the relationships derived for the base year are necessarily valid for forecast years. In recognition of this limitation, the Port Authority is engaged in a continued effort to update both the data input to the series of models sketched earlier and the modeling procedures themselves. Data for base year 1968 are now in the early stages of preparation.

ACKNOWLEDGMENT

The work described in this paper was developed jointly by The Port of New York Authority and the Traffic Research Group of Peat, Marwick, Livingston and Company, where Mr. Zupan was employed at the time of this work.

REFERENCES

1. Highway Traffic Estimation. Eno Foundation for Traffic Control, Saugatuck, Conn., 1956.

2. Moore, Edward F. The Shortest Path Through a Maze. International Symposium on the Theory of Switching, 1957.
3. Smock, Robert. An Iterative Assignment Approach to Capacity Restraint on Arterial Networks. HRB Bull. 347, 1962, pp. 60-66.
4. Traffic Assignment Manual. Bureau of Public Roads, U. S. Department of Commerce, 1964.
5. Schneider, Morton. Direct Estimation of Traffic Volume at a Point. Tri-State Interim Technical Report 4019-1320.
6. Brown, Geoffrey J. H., and Woehrle, Albert H., Jr. Program, Inputs and Calibration of the Direct Traffic Estimation Method. Highway Research Record 250, 1968, pp. 18-24.
7. Cherniack, Nathan. Measuring the Potential Traffic of a Proposed Vehicular Crossing. American Society of Civil Engineers, Transactions, Vol. 106, 1941, p. 520.

A Technique to Calibrate Choice Models

FRANCIS P. D. NAVIN, Alan M. Voorhees and Associates, Inc., and
GORDON W. SCHULTZ, R. H. Pratt Associates, Garrett Park, Maryland

The need for fast, efficient, and inexpensive techniques to calibrate transportation planning choice models led to the development of the methodology discussed in this paper. The calibrated models consist of stratified curves with a minimum of three variables to reflect conditions both at ends of a trip and on the competing transportation systems. The requirements of a stratified curve model are dealt with in considerable detail as well as their advantages to the model-builder and transportation planner. The calibration technique starts with an approximated set of curves and then attempts, by relaxation of one variable at a time, to fit an unknown set of curves in a rational manner. The first adjustment obtains the correct number of trip productions into the system as well as the number of attractions out of the system by traffic zones within certain characteristic strata. The curves regulating trip production and attraction are then adjusted to obtain the correct trip length over the model system variables. The final check on the trips is the origin-destination distribution by spider network assignments or other indirect techniques. This technique was used in Baltimore, Columbus, and Detroit during 1969, and the major results are presented in figures and tables. The model and technique are particularly suitable for computer application to large studies.

•PLANNING for future transportation facilities is dependent on good predictive models. Large sums of money and effort are expended in the calibration of these models. The calibration of the modal-choice model (transit-auto) is often critical to the configuration of the future mass transportation systems. Auto-occupancy models are also choice models that will divide highway trips between auto driver and passenger. This paper explains a simple technique to balance choice models that are constructed using stratified curves. The technique allows a systematic approach that will take a "guessed" set of curves and rapidly manipulate them toward the correct answer, thus saving much time and money. The method is sufficiently systematic to be computerized.

STRATIFIED CURVE MODELS

The stratifying of trips by purpose in the origin-destination (O-D) survey serves to group types of trips to minimize the variance of characteristics within the purpose categories and allows for a more meaningful analysis of travel patterns. (Many cities have successfully used stratified curves for modal-split models. Washington, Buffalo, and Seattle are a few examples.) Stratification may be carried one step further to independent variables to minimize variance within the group as well as to allow for nonlinear model relationships. The production and attraction variables are stratified in a meaningful manner to ensure a relative homogeneity within the strata.

Stratification has the additional advantage that the dependent variable and the trip-end and system variables may be visually examined and plotted to guard against inconsistent occurrences arising from insufficient data.

A choice model should be capable of explanation by the model-builder and understandable to the user. This usually precludes the use of a long complicated relationship

Paper sponsored by Committee on Origin and Destination and presented at the 49th Annual Meeting.

that duplicates the present but lends no rationale for the possible changes of the variables over time. Complex relationships are at best a risk because the interdependence of many variables may magnify variations. The variables for a choice model should describe

1. The trip-maker or the person making the choice,
2. The trip-maker's origin and destination,
3. The relative merits of the modes of transportation available, and
4. The trip purpose.

The trip-maker's characteristics and his origin are often interdependent, particularly in small and medium-sized cities. The trip-maker and his origin may be adequately described by a single variable such as income or residential density or car ownership.

The attraction end variable has some limitations placed on it simply because of the criteria of splitting travel between transit (public transportation) and the automobile (auto driver and auto passenger) and because the transit is usually aimed at intensive land-use areas such as the central business district (CBD). This implies that the attraction end variable should be selected to reflect intensive land-use areas. Some measures of the intensity of land use are employment density and parking cost.

The most critical variable to be selected is the variable that reflects the competitive position of the modes available for the journey. The competition of the systems may be stated as the ratio or difference (1) of two modal characteristics such as time or money or as some relative measure of the two. The waiting time and the time actually riding can be combined in some equitable manner to measure the apparent efficiencies of both systems (2, 3, 4).

TECHNIQUE

The technique is essentially a process of curve-fitting in a number of dimensions by the successive relaxation of one variable (relaxed parameter) until the estimated relationships approximate the unknown true relationship. A typical modal-choice surface (transit-auto) for a production-attraction pair and for two system variables S1 and S2 is shown in Figure 1. The usual view of the surface is the curve set A-E through D-H on the plane defined by X-S1; the parameter in this instance is the system variable S2. Similar curves may be projected on the plane Y-S2 with S1 and a parameter. The equation for the surface shown in Figure 1 has the form

$$Y = (X_1, X_2, \ldots X_n, S_1, S_2, \ldots S_n)$$

where

Y = percentage of trips taking choice A (area under the surface),
X_i = zonal production or attraction variables, and
S_j = a characteristic of the system or systems that connect the production and attraction zones.

The method assumes certain surface conditions, which for modal-choice models are usually self-evident and may be summarized as follows:

1. No discontinuity will develop on the surface.
2. The dependent variable usually will either monotonically increase or monotonically decrease with respect to independent variables; i.e., very few surface combinations will have valleys (5).

CORRECTION FOR NUMBER OF PREDICTED TRIPS

The starting point for the technique is a set of estimation curves and is a first attempt at approximating the true curve set. The first set of stratified curves may be any assumed approximation to the true unknown curves. If the curves closely approximate the true curves, then the time spent calibrating and refining the model is reduced.

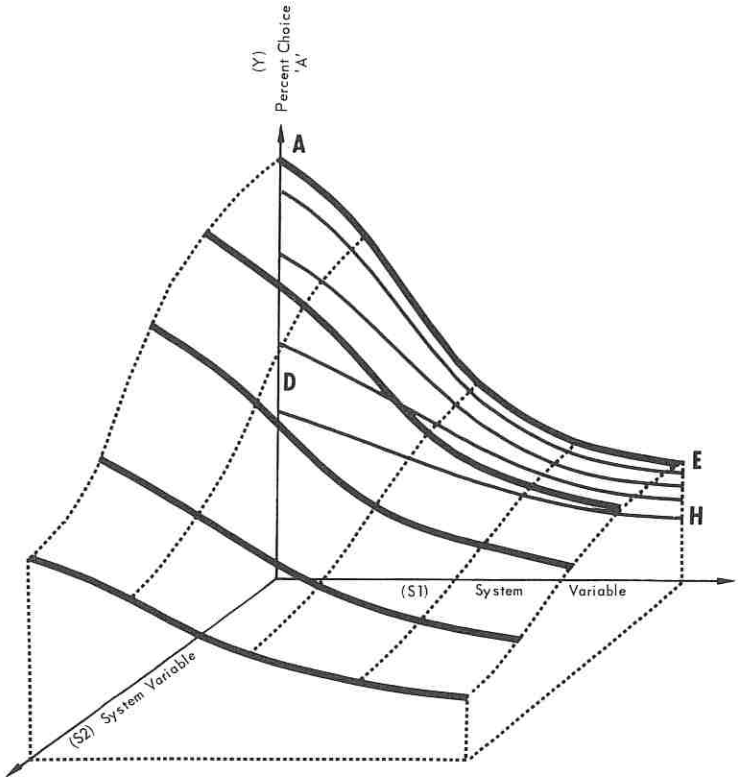


Figure 1. A choice model surface for a set of production and attraction variables and two system variables.

The first correction undertaken to make the estimation curves approach the unknown true curves is to adjust the number of trips while holding all other things constant. Assuming an even distribution of trips over the system variables, the percentage choice A is proportional to the number of trips. An average vertical shift correction for the entire curve along the dependent variable axis may be calculated in the areas adjacent to the known mean trip length along system variable S1. The average shift to be added to each point along the curve may be obtained from the equation

$$S_{ij} = \left(\overline{Y}S1 \times \frac{To_{ij}}{Te_{ij}}\right) - \overline{Y}S1 \qquad (1)$$

where

$\overline{Y}S1$ = percent choice A at mean trip length of the To trip-length frequency along S1,
To_{ij} = observed trips of choice A associated with the data set (ij), and
Te_{ij} = estimated trips of choice A associated with the data set (ij).

If two variables, income and employment density, are stratified into four classes each and if $\delta_{ij} = 0$ for each class combination, then the trips from all income classes would be predicted correctly. The attractions to all four employment density groups would also be correct. The number of surfaces associated with these variables would be 4 × 4 = 16 (assuming S1 and S2 are continuous). The corrections at various levels of trip grouping are given in Table 1, which represents a simple n × n trip matrix. If

TABLE 1

TRIP CORRECTION RATIOS FOR n x n PRODUCTION-ATTRACTION VARIABLES

Production Variable Strata	Attraction Variable Strata j = 1	j = 2	j = n	Total
i = 1	$\frac{To_{11}}{Te_{11}}$	$\frac{To_{12}}{Te_{12}}$ . . .	$\frac{To_{1n}}{Te_{1n}}$	$\frac{\sum_j To_{ij}}{\sum_j Te_{ij}}$
i = 2		$\frac{To_{ij}}{Te_{ij}}$		$\frac{\sum_j To_{ij}}{\sum_j Te_{ij}}$
i = n	$\frac{To_{n1}}{Te_{n1}}$			
Total	$\frac{\sum_j To_{i1}}{\sum_j Te_{i1}}$		$\frac{\sum_i To_{ij}}{\sum_i Te_{ij}}$	$\frac{To}{Te}$

Note: To = observed trips.
Te = estimated trips.

it is not possible to obtain trip values for each cell directly, then cell values may be obtained by simply setting up an equivalence table with a district for each cell of the matrix of Table 1 and by adding the resulting district-to-district trip table. The example being used will have a 16 × 16 matrix; and if k and l represent production attraction districts respectively, and if districts are numbered so that district 1 = T, and district 2 = T, . . . and district 16 = T, then the number of district trips for each strata cell of Table 2 is given by the relationship

$$T_{kj} = \sum_k \sum_l D_{kl} \tag{2}$$

where

$k = i \times (4 - 3)$ to $i \times 4$,
$l = j + (0, 4, 8, 12)$, and
D_{kl} = the trips in district interchange kl.

If it is not possible to set up such a district/production-attraction zone combination, then an approximate correction factor may be obtained from the following formula:

$$\text{Trip correction (ij)} = \frac{\sum_i To_{ij} + \sum_j To_{ij}}{\sum_i Te_{ij} + \sum_j Te_{ij}} \tag{3}$$

The best method of making this vertical shift is to have a choice program that accumulates trips within each production-attraction strata cell as the trips are being split between choices. The foregoing change will cause a vertical shift of δij along the dependent variable axis, percent choice A. If the trips are evenly distributed or if the trip-length distribution in all other dimensions are known to be correct, then no further corrections are necessary.

CORRECTION TO TRIP-LENGTH FREQUENCY IN S1 DIMENSION

The correct trip length of the trips of choice A must now be established in the S1 dimension. The S1 variable trip-length-frequency correction involves the rotation of the estimated curve set (ij) (shown in Fig. 2) about a line perpendicular to the S1 axis. The trip frequency of percentage of total trips over the system variable S1 is plotted for observed and predicted trips as shown in Figure 2.

Assuming that the trip-number correction has had no impact on the estimated trip-length frequency, then the point of rotation is given at the intersection of the observed and estimated distribution. The amount of rotation is calculated by the ratio of the area under the observed and estimated distributions within a few selected ranges. The S1 trip-length correction factor is

$$\xi_k = \frac{Ao_k^{S1}}{Ae_k^{S1}} \tag{4}$$

where

Ao_k^{S1} = the area under the observed distribution about the point k along S1, and

Ae_k^{S1} = the area under the estimated distribution about the point k along S1.

The trip-length correction will cause the curves to rotate in such a way as to approach the known true trip-length frequency. The correction may be more than a simple rotation if the curves intersect at more than one point. The corrected curve now has percentage choice A values of

$$\hat{Y}^{(3)}_{\substack{(ij)\\k}} = \left[\hat{Y}^{(1)}_{\substack{(ij)\\k}} + S_{ij}\right] \xi_k \qquad (5)$$

This equation is applied to all the curve projections A-E through D-H shown in Figure 1. The entire surface series of curves may be corrected approximately by ξ_k values using a trip-length frequency for the entire population. A refinement may be introduced by stratifying S2 and doing trip-length corrections for each strata of S2. The corrections would be calculated in a similar manner and would be applied to each curve that forms part of the curve set. If detailed trip lengths for stratified S2 values are not available, then the corrections in the S2 dimension would be undertaken in a manner similar to the correction of S1. The corrected percent choice A would be of the form

$$\hat{Y}^{(4)}_{\substack{(ij)\\k,1}} = C_1 \hat{Y}^{(3)}_{\substack{(ij)\\k}} = C_1 E_k \left(\hat{Y}^{(1)}_{\substack{(ij)\\k}} + S_{ij}\right) \qquad (6)$$

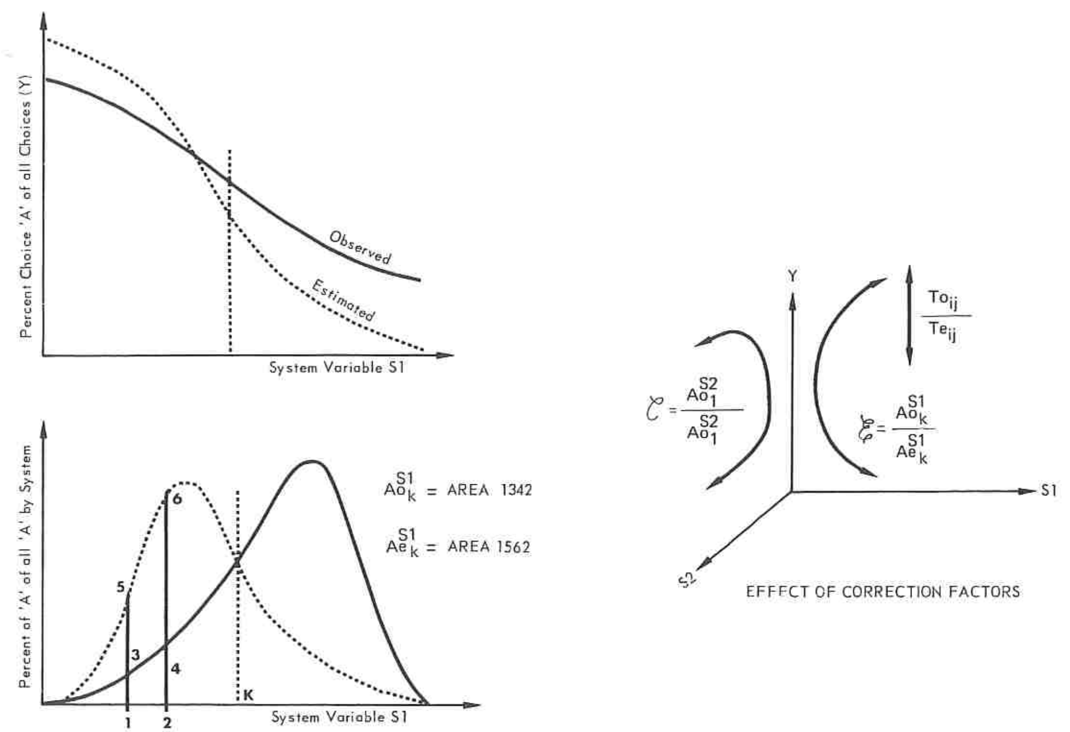


Figure 2. General model calibration procedure.

where

C_1 = the trip-length correction factor in region of the point 1 along axis S2, and
k = the point used for the S1 trip-length correction.

This then leads to the most general case of an n-dimensional relationship for the percent choice A.

$$Y\underset{k_1, k_2, \ldots k_n}{(X_1, X_2, \ldots X_n)} = \underset{k_1}{F_1} \times \underset{k_2}{F_2} \times \ldots \times \underset{k_n}{F_n} \left[Y(1)(X_1, X_2, \ldots X_n) + \underset{k_1}{S}_{X_1}, X_2 \ldots X_n \right] \quad (7)$$

Stratified curves lose most of their meaning beyond four independent variables. The method is such that if all the criteria are not satisfied, then new refined curves are used to estimate trips and the new trips and trip lengths are checked for acceptability. Because the trips are not evenly distributed over the system variables or over the variations of the data entering the model, the technique is an approximation that will approach the correct answer. The rate at which the model will calibrate is proportional to the number of independent variables used to explain (a) the choice model, (b) the regularity of the surface, and (c) the closeness of the first curve set to the true answer.

If it is possible to investigate each surface in all dimensions by parts, then a calibrated model may be obtained in a few iterations.

The technique just described has been applied at varying degrees of sophistication to five cities; the results of three studies are presented in the following section.

MODEL RESULTS

Stratified curve models and the technique just outlined were used recently for modal-split and auto-occupancy models in three cities. The Baltimore models investigated each set of curves individually and then collectively. The remaining two cities, Columbus, Ohio, and Detroit, Michigan, employed only the known number of trips associated with each pair of O-D variables as well as the total trips distributed over the system variables. The modal-split models (except NHB) employed the following form:

Percent of transit (Columbus): Median family income, employment density, and equivalent time difference (1);

Percent of transit (Baltimore): Median family income, parking cost, and equivalent time difference;

Percent of transit (Detroit): Median family income, employment density, travel cost difference, and equivalent time difference; and

Percent of auto driver (Baltimore and Detroit): Median family income, parking cost, and total highway travel time.

TABLE 2

TRIP CORRECTION RATIOS AND OBSERVED TRIPS

Production Variable (median family income)	Attraction Variable (parking cost)				Total
	j = 1	j = 2	j = 3	j = 4	
i = 1	1.055 9,181	1.196 238	1.360 279	0.714 152	1.058
i = 2	1.005 175,135	1.038 2,858	1.162 5,765	1.095 3,500	1.011
i = 3	1.034 1,040,691	1.985 24,990	0.966 32,012	1.000 22,552	1.032
i = 4	1.021 250,884	1.038 6,354	0.991 12,920	1.015 8,800	1.021
Total	1.029	1.073	0.985	1.011	1.028

Note: First row of figures is observed/estimated trips; second row is observed auto-driver work trips.

TABLE 3

MODEL CALIBRATION RESULTS FOR THREE CITIES

Purpose	Transit			Auto Driver	
	Baltimore	Columbus	Detroit	Baltimore	Detroit
Trips:					
Work	99,535	36,593	176,421	255,078	1,648,684
	100,050	36,389	176,533	252,037	1,613,553
Non-HB	3,643	17,409	32,901	37,167	1,315,590
	4,163	17,478	33,880	36,778	1,282,797
Other	13,516	16,164	116,346	87,481	2,587,463
	13,338	16,705	122,251	87,717	2,610,144
School	30,904		125,734	4,073	110,882
	30,905		126,605	5,695	112,474
Misc.		20,201			5,662,609
		20,763			5,618,968
Total	147,598	90,367	450,426	384,699	
	148,457	91,334	459,169	382,227	
CBD attractions:					
Work	32,253	19,747	48,826	24,768	67,735
	31,938	19,951	47,234	25,915	66,673
Non-HB	938	7,575	6,547	1,705	173,961
	919	7,400	7,424	1,760	160,851
Other	4,719	4,041	44,411	3,321	44,884
	4,392	4,132	41,570	3,652	45,028
School	937		1,362	187	3,992
	949		3,019	163	3,880
Misc.	38,847	11,146	101,136	29,981	290,572
	38,198	10,671	99,248	31,490	276,432
Total		42,509			
		42,154			
R^2 interchange:					
Work	0.42	0.66	NA	0.72	NA
Non-HB	0.13	0.38	NA	0.25	NA
Other	0.21	0.15	NA	0.64	NA
School	0.65		NA	0.10	NA
Misc.		0.63			

Note: Percent samples were: Baltimore, 5 percent; Columbus, 25 percent; Detroit, 4 percent. The first row of figures is the observed trips; the second, the estimated trips. R^2 = (explained variation)^2/(total variation)^2.

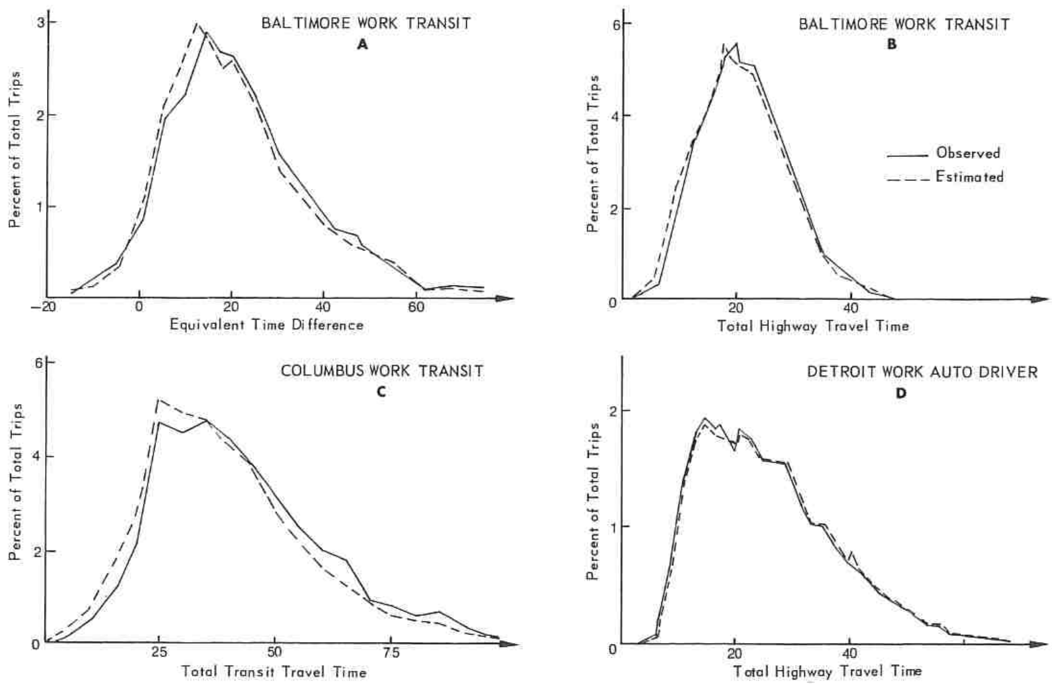


Figure 3. Trip-length-frequency comparison.

The equations all have a production variable, an attraction variable, and at least one continuous system variable. All variables were stratified except equivalent time difference and total highway travel time, which were continuous variables.

A few of the results obtained from these curves are given in Tables 2 and 3 and shown graphically in Figure 3. The results indicate the refinement that may be obtained in a model even when only a few variables are selected. The predictive accuracy of the Detroit work-auto-driver model by strata is given in Table 2. The trips are based on a 3 percent sample; therefore, the number of samples range from a low of five to a high of over 31,000. The precision is in line with the number of samples; the stratum with the largest number of samples varies from the observed by only 3.4 percent. The future use of the model has reliable curves to predict the total trips, provided the range of the variables is not exceeded.

This stratum accuracy check (a trip correction factor) established the number of trips between areas with certain characteristics and, therefore, gives a certain assurance that productions and attractions at the zonal level are approximately correct.

The next factor to investigate is the trip-length distribution along the system variable. The Baltimore work-transit trips provide an excellent example of a quick and inexpensive check on trip distribution.

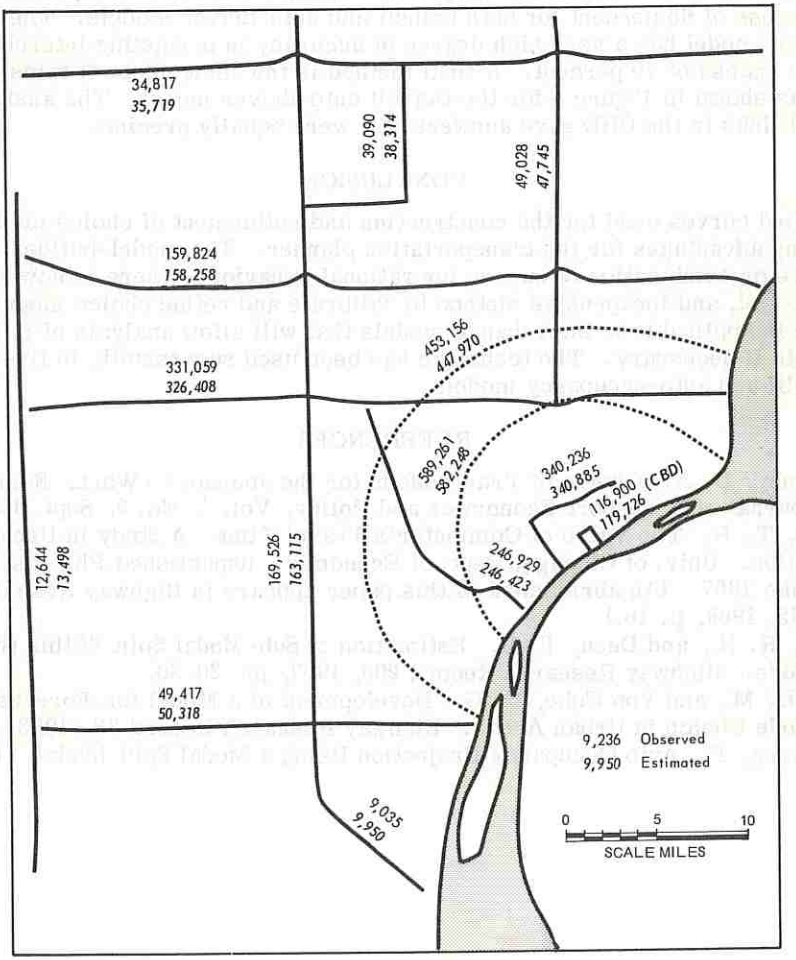


Figure 4. Detroit spider assignment of work-auto-driver trips.

The predicted trips were distributed over the model system variable; the result of this is shown in Figure 3A. A simple check on the correct trip O-D pair is a trip-length distribution over a variable that is not network dependent. The highway network (total highway travel time) was selected because it was less network dependent than transit network and because a spider network or centroid-to-centroid distance matrix was not available. The estimated Baltimore transit trips fit the distributions well for both variables, and the averages have approximately the same precision. The estimated Columbus transit trips distributed over the total transit travel time (Fig. 3C) coincide quite well with the observed distribution. The work-auto-driver trips distributed over the total highway travel time (Fig. 3D) gave an identical fit of the observed and estimated trips. The total hours of auto travel for work agreed within 0.5 percent. The trip-length distributions of Figures 3A and 3D show that a rotational correction is not required in either case; the transit distribution in Figure 3C indicates the transit trips are well distributed over the transit travel time. The distribution shown in Figure 3B is independent of the transit system and may approach a network-independent variable; this indicates the accuracy of the O-D pair selection.

The final check on the choice models is the correct O-D pair selection for each trip. This check may take many forms depending on what facilities and data are available. The first check may be distribution of trips over a network-independent variable such as centroid-to-centroid distance. A second test may be the interchange R^2 values as given in Table 3. The explained agreement ranges from a low of about 10 percent to a high in excess of 65 percent for both transit and auto-driver models. The Baltimore auto-driver model has a very high degree of accuracy in predicting interchange movements, in excess of 70 percent. A final method is the assignment of trips to a spider network as shown in Figure 4 for the Detroit auto-driver model. The assignment for individual links in the CBD gave answers that were equally precise.

CONCLUSION

Stratified curves used for the construction and refinement of choice models have a great many advantages for the transportation planner. The model-builder may inspect each curve or combination of curves for rational behavior. There is now available a practical, fast, and inexpensive method to calibrate and refine choice models. The technique is applicable to interchange models that will allow analysis of interchange movements if necessary. The technique has been used successfully in five cities for modal-split and auto-occupancy models.

REFERENCES

1. Quarmby, D. A. Choice of Travel Mode for the Journey to Work: Some Findings. Journal of Transport Economics and Policy, Vol. 1, No. 3, Sept. 1967.
2. Lisco, T. E. The Value of Commuter's Travel Time: A Study in Urban Transportation. Univ. of Chicago, Dept. of Economics, unpublished PhD dissertation, June 1967. (An abridgment of this paper appears in Highway Research Record 245, 1969, p. 16.)
3. Pratt, R. H., and Deen, T. B. Estimation of Sub-Modal Split Within the Transit Mode. Highway Research Record 205, 1967, pp. 20-30.
4. Hill, D. M., and Von Cube, H. G. Development of a Model for Forecasting Travel Mode Choice in Urban Areas. Highway Research Record 38, 1963, pp. 78-96.
5. Spielberg, F. Auto Occupancy Projection Using a Modal Split Model. Unpublished.

Evaluation of Bias in License Plate Traffic Survey Response

HOWARD McCANN and GARY MARING, Bureau of Public Roads,
Federal Highway Administration, U.S. Department of Transportation

This report presents the results of a second trial of license plate traffic survey procedures. The results of the initial trial were reported in Highway Research Record 297. The second survey was primarily designed to evaluate bias in response to mail questionnaires. Trip data from a license plate survey were compared with trip data from a conventional roadside interview. The study was conducted on Interstate 70 near Junction City, Kansas. The State Highway Commission of Kansas conducted conventional roadside interviews on August 12-15, 1968. The license plate traffic survey was conducted on August 27-28, 1968. This report provides background information, describes operating procedures during the survey, presents results of the survey, and gives conclusions and recommendations.

•AN INITIAL TRIAL of license plate traffic survey procedures was conducted in Massachusetts in January 1968. The results of the survey are reported in Highway Research Record 297. The paper is entitled "License Plate Traffic Survey."

In summary, vehicle license plates were recorded at highway stations, and questionnaires subsequently were mailed to vehicle owners requesting O-D information. Owners' names and addresses were determined by a computer search of the vehicle registration file. The high response rate of 60.3 percent indicated the motorists' willingness to respond to this type of survey. A limited analysis of the distribution of distance from registered address to highway station revealed no apparent difference between the respondents and nonrespondents.

Although the 60.3 percent response rate that was obtained during the initial survey was high, a bias in considering that nonrespondents have the same characteristics as respondents could produce considerable error in the survey results. A second trial of the survey procedures was designed specifically to evaluate bias in the response. This was accomplished by comparing trip data from a license plate survey with that from conventional roadside interviews.

The license plate traffic survey was conducted on Interstate 70, in Kansas, shortly after the State Highway Commission conducted roadside interviews. The state has performed a series of roadside interviews on this highway to measure traffic diversion and generation (1).

Vehicle license plates were recorded on Interstate 70 at a site about 3 miles west of Junction City, Kansas (Fig. 1).

Westbound traffic was surveyed on Tuesday, August 27; and eastbound traffic was surveyed on Wednesday, August 28, 1968. Roadside interviews had been conducted at this station from Monday, August 12, through Thursday, August 15, 1968. Comparisons were made for the same hours in both surveys, although the days of the week were different.

Paper sponsored by Committee on Origin and Destination and presented at the 49th Annual Meeting.

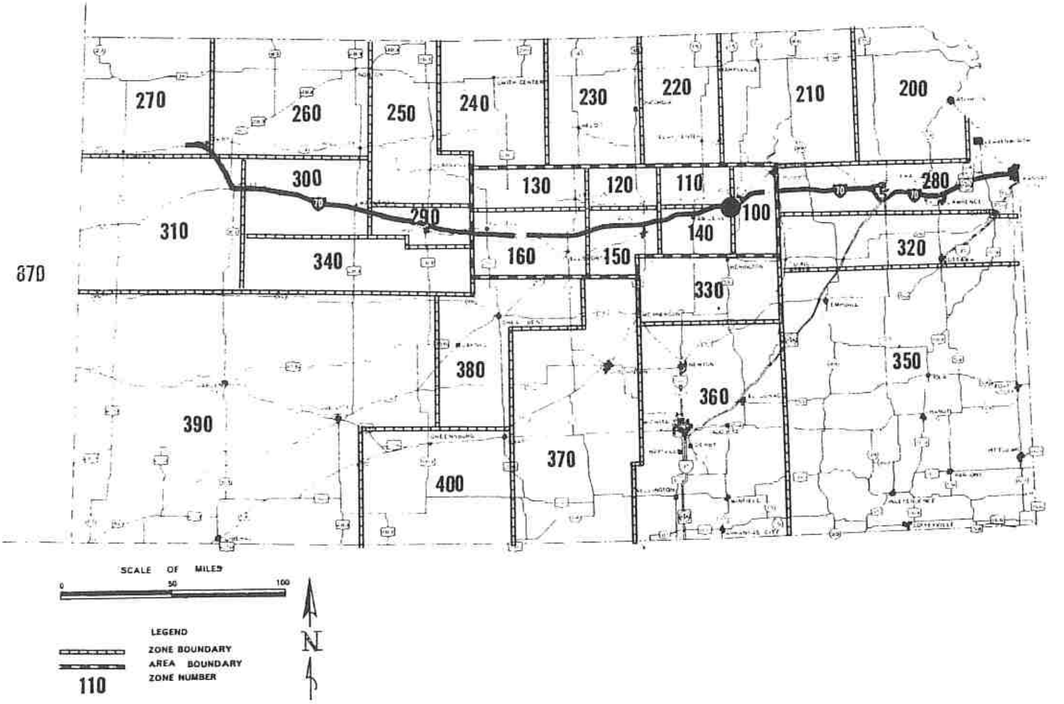


Figure 1. Kansas map showing survey location and traffic zones used for survey comparisons.

PROCEDURES

The survey procedures were similar to those followed in the Massachusetts survey. Two innovations were used for the Kansas survey:

1. The cameras were equipped with electronic switches that exposed one frame of film as a vehicle passed a road tube. The switches were developed by BPR and cost about $30.00 each (Fig. 2).

2. The vehicle registration data were printed on the postal questionnaires by the computer.

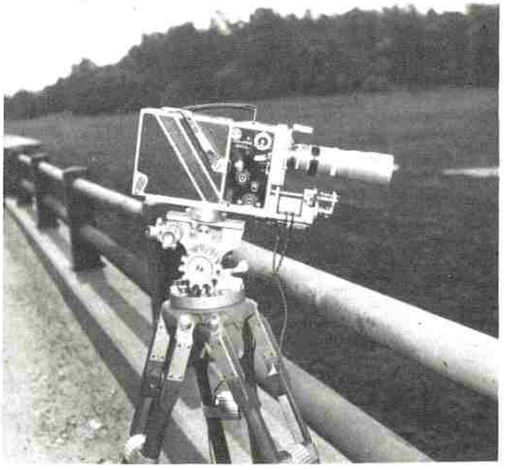


Figure 2. Camera with electronic switch.

Description of Computer Search

Data processing cards were punched to indicate the license numbers of the cars photographed on the highway. Kodak 310 film readers were used to enlarge the film for reading by the keypunch personnel.

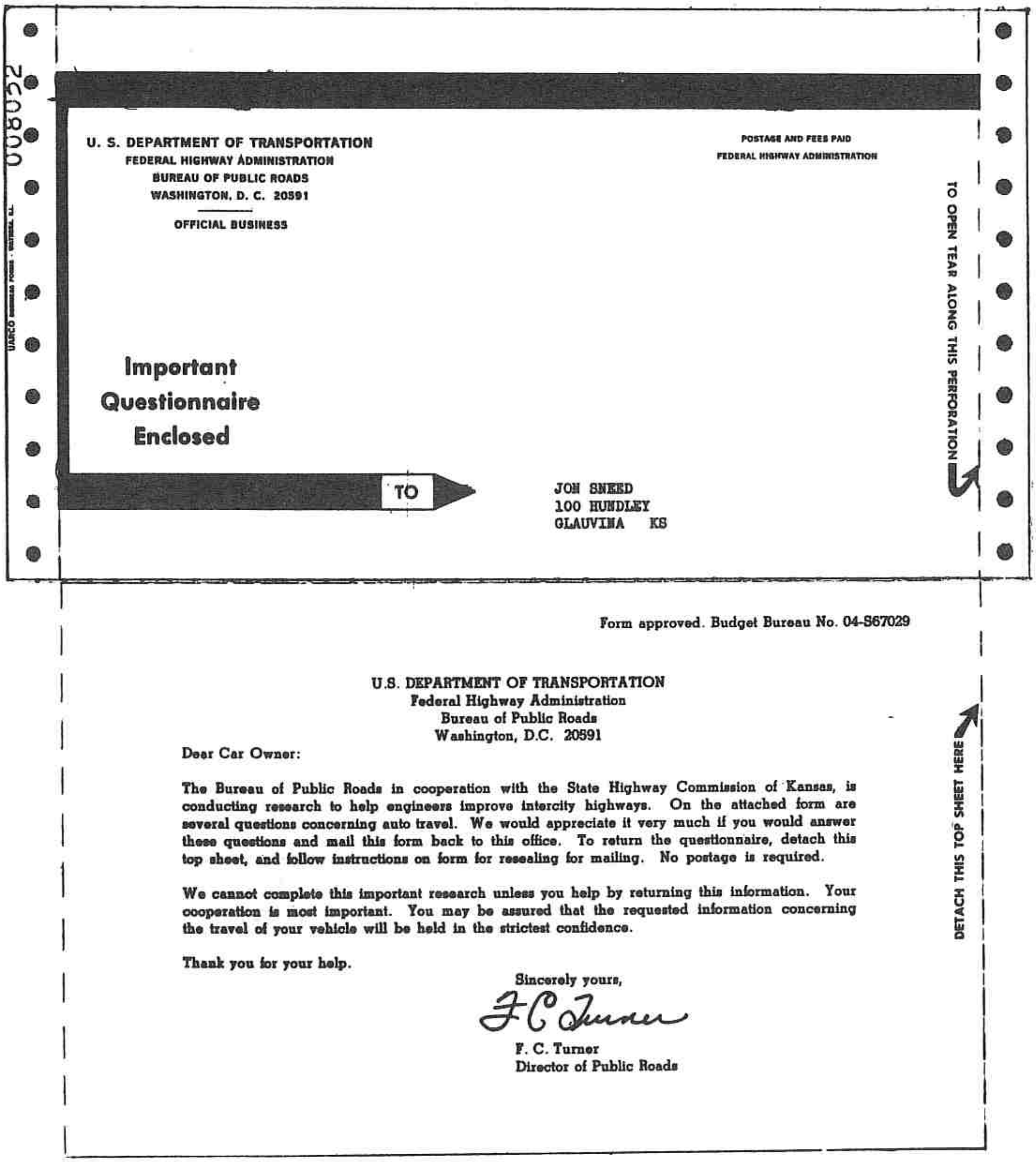
A computer search determined the vehicle owners' names and addresses. The Kansas Motor Vehicle Department has an IBM 360 computer with the file contained on a direct-access device.

When a vehicle license number was located in the file, the computers printed the necessary data for postal addressing and vehicle identification on the questionnaires.

The registration file did not contain zip codes, thus prohibiting the use of bulk postal rates. The cost of mailing questionnaires was the first-class rate of 6 cents.

Questionnaire Design

The questionnaires were printed on marginally punched paper for use on the computer (Figs. 3 and 4). The computer printed the survey date, location, owner's name and address, and vehicle description (consisting of year, manufacturer, and license plate number).


U. S. DEPARTMENT OF TRANSPORTATION
FEDERAL HIGHWAY ADMINISTRATION
BUREAU OF PUBLIC ROADS
WASHINGTON, D. C. 20591

OFFICIAL BUSINESS

POSTAGE AND FEES PAID
FEDERAL HIGHWAY ADMINISTRATION

TO OPEN TEAR ALONG THIS PERFORATION

Important Questionnaire Enclosed

TO

JOH SNEED
100 HUNDLEY
GLAUVINA KS

Form approved. Budget Bureau No. 04-S67029

U.S. DEPARTMENT OF TRANSPORTATION
Federal Highway Administration
Bureau of Public Roads
Washington, D.C. 20591

Dear Car Owner:

The Bureau of Public Roads in cooperation with the State Highway Commission of Kansas, is conducting research to help engineers improve intercity highways. On the attached form are several questions concerning auto travel. We would appreciate it very much if you would answer these questions and mail this form back to this office. To return the questionnaire, detach this top sheet, and follow instructions on form for resealing for mailing. No postage is required.

We cannot complete this important research unless you help by returning this information. Your cooperation is most important. You may be assured that the requested information concerning the travel of your vehicle will be held in the strictest confidence.

Thank you for your help.

Sincerely yours,

F C Turner

F. C. Turner
Director of Public Roads

DETACH THIS TOP SHEET HERE

Figure 3. Postal questionnaire.

QUESTIONNAIRE

Your car, described at the bottom of this page, was observed on

WEDNESDAY, AUG. 28, GOING EAST ON I-70 NEAR JUNCTION CITY, KANSAS

OFFICE USE ONLY

Listed below are several questions concerning that trip. Please answer these questions and return this form.

1. Did this trip start at your home address on the above date? Yes ☐, No ☐
   If no, where did the trip start? City or place ______
   (check one) In Rural area ☐ or Urban area ☐
   State ______

2. Did this trip end at your home address on the above date? Yes ☐, No ☐
   If no, where did the trip end? City or place ______
   (check one) In Rural area ☐ or Urban area ☐
   State ______

3. What was the destination of this overall trip? (This is normally the farthest point to which the vehicle was driven from the home address).
   City or place ______
   (check one) In Rural area ☐ or Urban area ☐
   State ______

(QUESTIONS CONTINUE ON NEXT PAGE)

VEHICLE DESCRIPTION	OWNER IDENTIFICATION
66 CHEV SN000000	JON SNEED 100 HUNDLEY GLAUVINA KS

OFFICE USE ONLY

4. What was the approximate length of the trip from the home to the destination? ______ miles.

5. What was the purpose of the trip?
   - Earning a living ☐
   - Family business ☐
   - Social, recreational ☐
   - Vacation ☐
   - Shopping ☐
   - Other ______ ☐

6. How many occupants, including the driver, were in the vehicle? ______

7. What were the ages of the occupants?

Age group:	Number
Under 18	______
18 - 44	______
45 - 65	______
Over 65	______

8. Is the above vehicle kept at the registered address? (check one) Yes ☐ No ☐
   If no, where is the vehicle kept?
   Street ______, City or place ______, State ______

9. Is the vehicle kept in a rural area, or within an urban area (within city limits or suburb)? Rural area ☐
   Urban area (within city limits or suburb) ☐

Errors in recording license plates do occur. If this form was sent to you through error, please check here ☐ and return the form.

THANK YOU FOR YOUR COOPERATION

Figure 4. Postal questionnaire.

Each questionnaire, measuring $5\frac{1}{2}$ by $9\frac{7}{8}$ in. consisted of four sheets of marginally punched paper. The first sheet contained the owner's name and address on one side and a message explaining the survey on the other side. The second sheet included the Bureau of Public Roads address for returning the questionnaire. The third and fourth sheets contained questions to be answered by the respondent. To return the questionnaire, the first sheet was detached; and the second and fourth sheets were joined with pressure-sensitive tape.

The questionnaires cost $5\frac{1}{2}$ cents each. For a large order of 100,000 or more, the costs would be reduced to about 2 cents per questionnaire.

Questionnaire Mail-Out

A total of 1,815 questionnaires were mailed to Kansas vehicle owners on August 30, 2 days after the filming was completed. No follow-up procedures were used for this survey.

The 1,815 vehicles recorded represented a sample of the total Kansas vehicles passing the station. Intermittent rain fell during the 2 survey days, and all license plates could not be recorded.

Reliability of Response

Two procedures were used to test the data for bias. The first test was a comparison of the distributions of trip purpose, vehicle occupancy, and zone-to-zone trip movements obtained in the license plate survey and in the roadside interview survey. The second test was an analysis of response rate and trip length by age of vehicle owned.

The above comparisons were made only for vehicles registered in Kansas. The large number of vehicles recorded from other states precluded the mailing of questionnaires to all out-of-state vehicle owners for the purpose of this study.

RESULTS

The response rate was similar to that in the Massachusetts survey, with 69.1 percent of the questionnaires returned by October 14 (Table 1). This percentage was reduced to 63.5 when unusable responses were eliminated.

Table 2 gives a comparison of the trips reported between specific zones for the two surveys. Only the zonal interchanges with enough trips to be considered significant were included. The table gives a close comparison between the results of the two surveys.

TABLE 1

RESPONSE RATE FOR KANSAS VEHICLE OWNERS

Category	Number	Percent
Total sent out	1,815	100.0
Returned	1,254	69.1
Usable	1,152	63.5
Errors in recording plates, undelivered, etc.	102	5.6

TABLE 2

COMPARISON OF ZONE-TO-ZONE TRIPS REPORTED BY KANSAS VEHICLE OWNERS IN THE LICENSE PLATE SURVEY AND THE ROADSIDE INTERVIEWS[a]

Zonal Pairs	Approximate Distance Between Zones (miles)	Percent of Trips[b]	
		Roadside Interviews	License Plate Survey
100-140	24	25.5	23.6
100-150	50	13.3	11.1
140-280	117	4.0	4.3
150-280	143	6.3	7.2
280-870	550	6.9	5.8
280-370	220	2.3	3.3
280-380	220	2.4	1.1

[a]Zone locations are shown in Figure 1.
[b]All trips reported by Kansas vehicle owners.

TABLE 3

COMPARISON OF TRIP PURPOSE AND VEHICLE OCCUPANCY REPORTED BY KANSAS VEHICLE OWNERS IN THE LICENSE PLATE SURVEY AND THE ROADSIDE INTERVIEWS

Trip Purpose	Roadside Interviews		License Plate Survey	
	Percent of Trips	Vehicle Occupancy	Percent of Trips	Vehicle Occupancy
Work	35.9	1.47	39.4	1.44
Family business	15.0	2.05	15.9	2.58
Social-recreational	24.6	2.89	8.2	2.40
Vacation	17.0	3.31	13.7	2.95
Shopping	3.5	2.47	4.7	2.62
Other	4.0	2.48	18.1	2.00
Total (avg)	100.0	(2.29)	100.0	(2.06)

Table 3 gives a comparison of trip purpose and vehicle occupancy distributions from the license plate survey with those from the roadside interview survey. The percentage of trips and vehicle occupancy is similar for three of the purposes—work, vacation, and shopping. The differences for the other purposes—family business, social-recreation, and other—could be a result of the following factors: the survey may produce respondents with bias, and the trip purpose categories may not be understood by the motorists.

Considerable judgment must be used in evaluating the effect of the individual factors. If the first factor—the bias of respondents—were dominant, it is doubtful that any of the purpose groups would have compared closely. Three trip-purpose groups—work, vacation, and shopping—had similar percentages of trips and vehicle occupancy for the two surveys. Also, if the bias were dominant, it is unlikely that the two surveys would have produced comparable percentages of zone-to-zone trips.

TABLE 4

DISTRIBUTION OF VEHICLE AGE FOR RESPONDENTS AND ALL VEHICLES SAMPLED ON THE ROADWAY DURING THE LICENSE PLATE SURVEY

Age of Vehicle	Percentage of Vehicles	
	Respondents (percent)	Total Roadside Sample (percent)
1959 or earlier	7.9	10.0
1960	1.9	2.5
1961	2.6	2.7
1962	5.0	5.5
1963	9.8	9.5
1964	12.3	11.1
1965	9.8	9.4
1966	14.4	14.0
1967	20.6	18.6
1968	15.7	16.7
Total	100.0	100.0

Taking this evidence into account, it would be difficult to conclude that bias had a dominant effect. More likely, motorists simply do not understand the meanings of some trip-purpose terms. In a personal interview, the interviewer can assist in clarifying the meanings of the groups and thus guide the motorist to a proper answer.

Table 4 gives a comparison of the age of the respondents' vehicles with the age of all vehicles in the survey. Although the comparison was close, there appeared to be a slight tendency for the respondents to own newer vehicles. For example, 10.0 percent of all the cars in the survey were made in 1959 or earlier compared to 7.9 percent of the respondents' cars.

Trip lengths reported by the respondents showed that cars manufactured before 1963 had an average trip length of 166 miles, whereas 1963 through 1968 cars had an average trip length of 281 miles. This variance in trip length would indicate that the license plate survey would show slightly fewer short trips than the Kansas roadside survey since persons with older cars answered at a slightly lower rate than those with newer cars. This trend is given in Table 2. The shortest zone-to-zone movements were zones 100 to 140 and zones 100 to 150. The license plate survey showed a slightly lower percentage of trips in these categories than the Kansas roadside survey.

The approximate cost per returned, and usable, questionnaire was 42 cents. The breakdowns for this cost figure are shown below. The cost of 42 cents per interview is comparable to the cost of roadside interviews on high-volume, high-speed roads.

Computer search	\$0.02 per questionnaire mailed
Questionnaire	$0.05\frac{1}{2}$
Postage	0.06
Film	0.02
Keypunching	0.05
Camera rental	0.02
	$\$0.22\frac{1}{2}$ per questionnaire mailed

Return rate 63.5 percent

$$\text{Cost per returned questionnaire} = \frac{100}{63.5} \times 0.22\tfrac{1}{2} = 0.35\tfrac{1}{2}$$

$$\text{Postage for returned questionnaire} = 0.06$$

$$\text{Cost per returned questionnaire} \quad \$0.41\tfrac{1}{2}$$

CONCLUSIONS

It appears that a license plate traffic survey and a roadside interview survey produce comparable numbers of trips between zones. The two types of surveys were comparable in three trip-purpose groups—work, vacation, and shopping. The other trip-purpose groups—family business, social-recreation, and other—showed some variations. These variations can probably be primarily attributed to a misunderstanding of the meaning of these trip purposes.

The 63.5 percent questionnaire response rate of this survey was very similar to the 60.3 percent response rate of the Massachusetts survey. The consistency of survey results demonstrates the response rate that may be expected from license plate surveys. Higher response rates could undoubtedly be obtained with the use of reminders or follow-up questionnaires.

RECOMMENDATION

The trip-purpose question could probably be improved for a license plate survey by adding more purposes to the list and by omitting "other." This might have the effect, comparable to that of the personal interview, of guiding the respondent to a specific purpose group. Short definitions might also help to clarify certain trip purposes.

ACKNOWLEDGMENTS

Mr. John McNeal and Mr. L. A. Billings, State Highway Commission of Kansas, provided assistance in all phases of the survey.

Mr. Leon E. Litz, Chief of the Methods Branch, and Mr. Darel Trueblood, Chief, Planning and Research Division, Region 5, aided in the technical design of the survey and in discussing and resolving problems relative to the study.

Considerable credit is due to Mr. William T. Jordan, Planning and Research Engineer, and Mr. Ronald A. Shriver, Assistant Planning and Research Engineer, Kansas Division, for their efforts during the survey.

REFERENCE

1. Interstate 70 Usage Study, Junction City to Russell, 1962-1965. State Highway Commission of Kansas.

Deriving the Traffic Consequences of Airport Location Alternatives

P. M. PEARSON, Memorial University of Newfoundland, and
W. A. McLAUGHLIN, University of Waterloo

The primary objectives of the research reported in this paper were to structure the airport location process and to develop a methodology for deriving the traffic consequences of various airport location alternatives. A number of interconnected analyses were identified in the location procedure, including demand forecasting, constraint recognition, cost estimates, and airport location evaluation. A demand model based on systems engineering concepts was presented. Linear graph analysis was used to describe mathematically the travel volumes on each link of the intercity travel network. It was shown that by using the complementary travel pressure variable, the traffic consequences of various airport locations on the short-haul travel market could be derived. Finally, the results of the model were used to determine the user travel benefits associated with each of three Toronto airport location alternatives.

•RECENT STUDIES of airport development (1, 2, 3) have considered variables such as land costs, ground transportation costs, meteorological factors, and aircraft noise contours for optimum airport location. None of these studies, however, has related the demand for air travel to the airport location.

The location of the airport defines the ground transportation portion of the air trip in terms of travel costs and travel times. Considering both ends of a trip, ground travel can be in excess of 60 percent of the total air trip time for lengths of less than 300 miles (4). Furthermore, within the 100- to 400-mile trip lengths, air travel is in direct competition with other intercity modes. Any increase or decrease in the ground portion of the air trip can alter the existing intercity modal distribution as well as the total number of air travelers. In the short-haul distances, it is unrealistic to assume that the location of an airport has no effect on the demand for air travel.

It is the objective of this paper to present a method by which the traffic consequences of airport locations can be derived. The technique is based on systems theory, which requires that each of the individual components of a system be defined mathematically and that these components be interconnected to form a complete interdependent and analytical model of, in this example, an intercity travel network. The technique is applied to several Toronto International Airport locations, and the consequences on Toronto-Montreal and Toronto-Ottawa traffic are derived.

LOCATION ANALYSIS PROCEDURE

The basic objective that must be fulfilled by an airport system may be stated as minimizing the sum of the capital and operating cost of the airport terminal system and the ground transportation costs of passengers, consistent with satisfactory ground access times and the constraints imposed by navigational and safety requirements and those by human habitation (5).

Paper sponsored by Committee on Origin and Destination and presented at the 49th Annual Meeting.

Figure 1 shows the location process structured into a logical framework of interconnected analyses. The first phase requires a statement of future air travel demands, and this is a statement of need for new or expanded facilities. Methodologies of air traffic forecasting have been presented by a number of authors (6, 7, 8, 9, 10).

The next phase of the framework requires that the various costs be assessed for each location alternative. At this level of comparison, the major objective is to choose between different airport sites. General cost figures are required so that a decision can be made between broad classes of airport location solutions. Costs germane to the framework include

1. Overall construction costs, including support facilities such as new connecting roadways;
2. Operating costs, including salaries, overhead, and maintenance for the planned life of the project; and
3. Operating revenues for the project life.

These costs should be considered using an appropriate interest rate. Wohl (11) has suggested that the interest rate is incorporated easily by calculating the net present value. This reduces all future monies to present-day terms. The net present value is obtained from

$$NPV(C) = \sum_{0}^{t} \frac{C_k}{(1 + i)^k} - \sum_{0}^{t} \frac{R_k}{(1 + i)^k} \qquad (1)$$

where

t = number of years of the project;
NPV(C) = net present value of costs;
C = total costs occurring in year k;
i = interest rate; and
R = revenues occurring in year k.

The next phase of the framework (the main area of interest in this paper) requires the evaluation of the total number of air passengers, their origins and destinations, and the level of ground transport service associated with each airport location alternative. The ground travel costs and times can be traded off against the total airport costs.

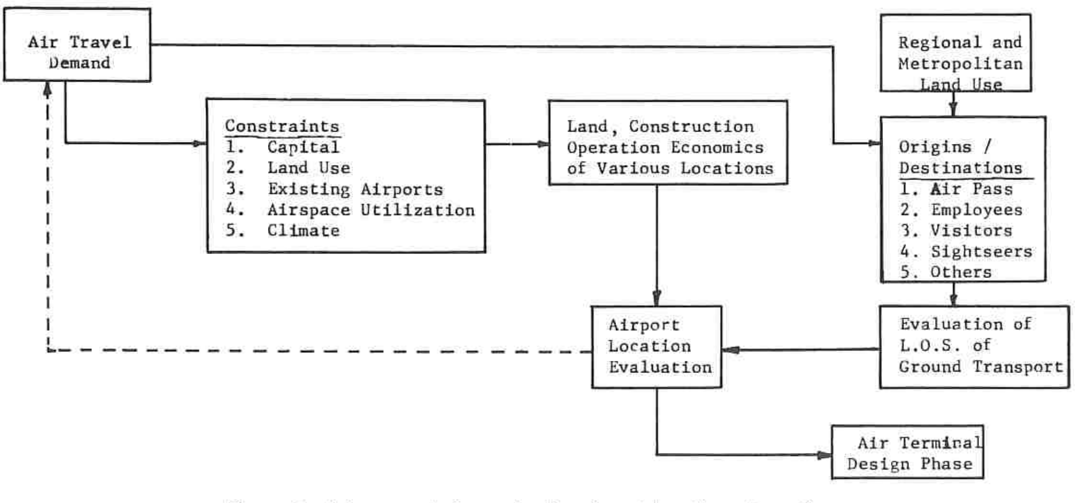


Figure 1. A framework for evaluating airport location alternatives.

In the evaluation phase, the optional project can be defined by

$$NPV(B - C) = NPV(C) - \sum_{o}^{t} \frac{B_k}{(1 + i)^k} \tag{2}$$

where B is the benefits occurring in year k.

AIR TRAVEL DEMAND CONSIDERATIONS

In most airport location studies, the traffic volumes are assumed constant. In other words, only the travel costs and times vary with airport location. In the short-haul distances, however, air travel is in competition with the highway and rail modes. Any variation within the ground portion of the trip can result in an increased or decreased number of air passengers.

Figure 2, for example, shows the traffic volumes for two airport location alternatives. For the existing intercity system, the equilibrium volume is V_{e1}, which is the sum of the air (V_{a1}), rail (V_{r1}), and highway (V_{h1}) volumes. In relation to the system prices P, the total travel benefits from the proposed airport locations are

$$\tfrac{1}{2}(P_{e1} - P_{e2})(V_{e1} + V_{e2}) = \tfrac{1}{2}(P_{a1} - P_{a2})(V_{a1} + V_{a2}) + \tfrac{1}{2}(P_{r1} - P_{r2})(V_{r1} + V_{r2}) + \tfrac{1}{2}(P_{h1} - P_{h2})(V_{h1} - V_{h2}) \tag{3}$$

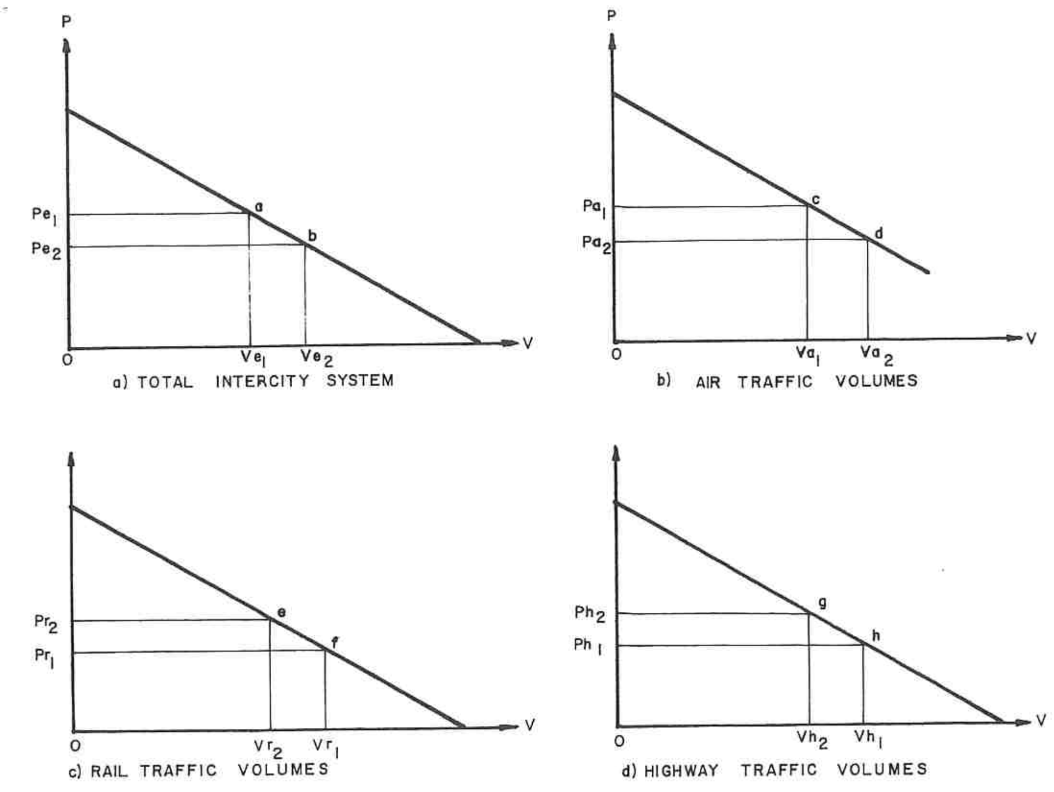


Figure 2. Total system benefits resulting from construction of a new airport.

In terms of the total system travel, the increase in benefits can be derived by determining the system equilibrium prices in terms of the weighted prices of the three modes. The prices weighted by traffic volumes are

$$P_{e1} = \frac{V_{a1}}{V_{e1}} P_{a1} + \frac{V_{r1}}{V_{e1}} P_{r1} + \frac{V_{h1}}{V_{e1}} P_{h1}$$

and

$$P_{e2} = \frac{V_{a2}}{V_{e2}} P_{a2} + \frac{V_{r2}}{V_{e2}} P_{r2} + \frac{V_{h2}}{V_{e2}} \tag{4}$$

There are diseconomies associated with inaccurate demand estimates. If the demand is overestimated, capital cannot be recovered during the facility's service life or, in the case of staged construction, during a planned development stage. This results in a loss of investment opportunity. If the demand is underestimated, the planned facility will not perform adequately. If an early retirement results, invested capital will not be recovered. If no further investment occurs, and the facility is forced to operate under unsatisfactory conditions, losses to the economy because of delays or lost traffic will be incurred. These diseconomies are shown in Figure 3a.

Figure 3b shows the diseconomies associated with underestimates of demand with project reinvestment. A project phased into the existing system was chosen considering the benefits and costs reduced to time $t = 0$. The project exhibited the following cost characteristics:

1. An initial investment of C_1 dollars for stage I; and
2. An investment of C_2 dollars at year n for stage II.

The value of this investment was $C_2/(1 + i)^n$ at $t = 0$. The project's performance, however, became unsatisfactory at time k ($k > n$). Stage II then was constructed at a cost C_2. Finally, an additional investment of C_3 was required at time r.

A number of costs were not considered by the decision-maker at time $t = 0$. The reduced value of these costs are as follows:

1. $C_2\left[1/(1 + i)^k - 1/(1 + i)^n\right]$, which is the additional cost resulting from the premature construction of stage II.
2. $C_3\left[1/(1 + i)^r\right]$, which is the additional cost caused by the investment in year r.

It is recognized that the benefits resulting from the unanticipated traffic volumes also were not considered by the planner. In fact, the actual history of the project may represent the "best" solution. Two distinct diseconomies exist, however, and these are

1. The additional capital was not considered in the planning process and, therefore, the project incurred an additional cost; and
2. Had all costs (and benefits) been included, there is a distinct possibility that an alternative project would have been chosen.

A MODEL OF INTERCITY TRAVEL DEMAND

Systems engineering techniques were applied to the Toronto-Ottawa and Toronto-Montreal intercity travel system. With the application of linear graph analysis, it was possible to derive the traffic consequences of several Toronto airport locations.

Linear graph analysis requires individual components to be modeled separately in terms of complementary pressure and flow variables. The imposition of their interconnection pattern then yields a model for the entire system. The procedure is analytic in form and theory and provides a consistent and rigorous approach for modeling systems. Furthermore, these techniques have been applied to socioeconomic systems, including traffic networks (12, 13, 14).

Linear graph analysis, as presented in this paper, was used to develop a set of linear equations that characterize the flow on all links of the intercity travel network (including the airport access). Because every link on the system is described math-

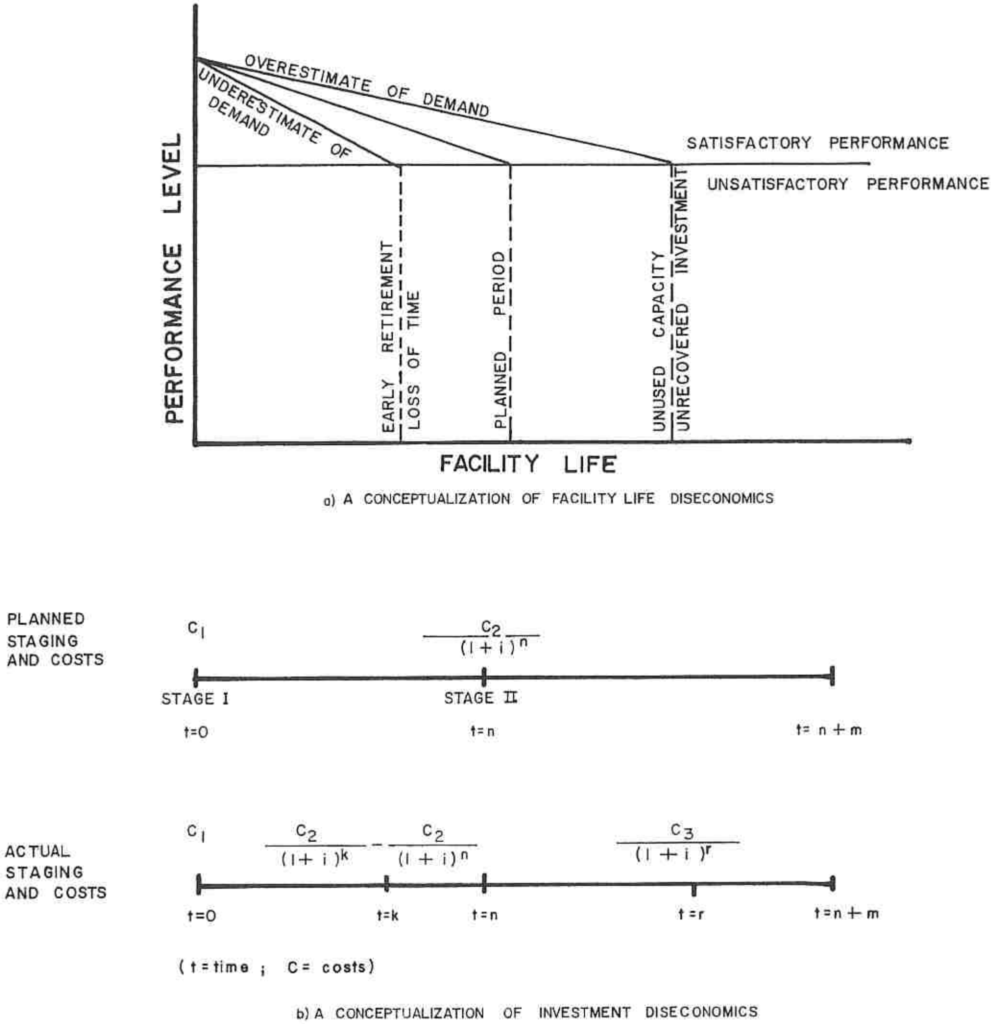


Figure 3. Diseconomies resulting from inaccurate estimates of travel demand.

ematically, the equilibrium demand components of generation, distribution, model distribution, and assignment are completed simultaneously.

The model development is presented under the following subheadings:

1. System identification by purpose and function and component choice;
2. Component measurement;
3. Components' terminal equations;
4. System graph;
5. System equations; and
6. Model results.

System Identification and Component Choice

The primary purpose of the model was to simulate the demand for intercity travel by mode associated with several Toronto airport locations. Because of data limitations, the study was restricted to annual business travel. The components included the Toronto

airport region generators; access links to the airport, rail, and highway terminals; the intercity routes to Montreal and Ottawa; and measures of the Montreal and Ottawa destination attractions.

Measurement on Components

Linear graph analysis requires that the following requirements be met:

1. The individual system components must be quantitatively describable by two fundamental variables. These variables are a y or flow variable and a complementary x or pressure variable that causes flow.
2. The components are connected at their ends (vertices) to yield a model for the entire system. The interconnected model must satisfy the two generalized Kirchoff laws. The first law states that the algebraic sum of all flows (y) at a vertex is zero. The second law states that the algebraic sum of all pressures around any closed loop of the system must be zero.
3. The flow and pressure variables must be related by a linear or nonlinear function.

The y variable for the intercity travel network is person-trips per year. This satisfies the first Kirchoff law and eliminates the necessity of modeling for storage within the system. That is, all business travelers are assumed to return to their origin over the yearly period.

The x variable is postulated to be a value measure used by the travelers in making a trip and a choice of mode. It is analogous to the portion of the travel potential of an origin that is used as a trip is made and thus is the pressure that caused flow. The x variable is not a measure of the total perceived value of making trips but rather the measurable perceived total cost of making the trip.

The reasoning for the above postulates is as follows:

1. If it is believed that the making of a trip and the choice of mode can be simulated with a reasonable degree of accuracy, then it follows that there is some underlying process made by the traveler in making such a choice.
2. The traveler will act as a free agent and attempt to optimize his degree of satisfaction.
3. Relating points 1 and 2 to a value measurement used in travel allows the origin pressure to dissipate as the trip is made, thus satisfying the second Kirchoff law.

The application of the second Kirchoff postulate to traffic networks as described in this paper requires an assumption, which is: The perceived cost (pressure) of a trip from any particular origin to all destinations is a constant. The reasons for this assumption are as follows:

1. Each origin is modeled with its own unique travel pressure.
2. The origin and travel links form a closed loop with each destination.

Terminal Equations

The origin areas can be characterized as a known flow driver of the form

$$Yi = y \tag{5}$$

where

Yi = the flow from origin i in annual business trips; and
y = a specified value taken from actual data.

The model is built from the existing system data. It then is solved for the complementary pressure variable, and the complementary model is constructed. Then changes can be made in the system to determine the changes in traffic volumes resulting from the implementation of a new airport facility.

The pressure variable was postulated to be of the form

$$X = A(I.P.) + B \tag{6}$$

where

X = the travel potential in cost per year;
A, B = regression constants; and
I.P. = origin area income, population cross product.

This relationship has been verified for the Canadian domestic airway system (14).

The route components have terminal equations of the form

$$X_{ij} = R(y) \times Y_{ij} \tag{7}$$

where

X_{ij} = the perceived value or cost used by the business traveler in crossing the link;
Y_{ij} = the flow in persons per year on link ij; and
$R(y)$ = the resistance function.

The resistance function is of the form

$$R(y) = C(y) + T(y) \tag{8}$$

where

$C(y)$ = the cost in cents to cross a link; and
$T(y)$ = the time, including delay, to cross a link translated to cents per person.

Calibration of the model (14) showed that the time translation constant was 10 cents per minute for air travelers, 6.5 cents per minute for auto travelers, and 15.0 cents per minute for rail travelers. The high value for rail reflects the high rail time and inconvenience perceived by the aggregate of travelers.

The access links included measures of terminal processing times. The egress links included the costs and times associated with overnight stops required for the various modes. Furthermore, the egress links included measures of modal competition (13).

The terminal equations of the destination cities were expressed as

$$Y_k = A_k X_k \tag{9}$$

where

Y_k = the number of business trips per year arriving at destination k;
A_k = the attraction of city k; and
X_k = the cost used across city k.

The attraction measures were based on a study by Air Canada (15). The function was of the form

$$A_k = \varphi\left[\beta_S\left(e_{sk}/e_{s,avg}\right) + \beta_H\left(e_{Hk}/e_{H,avg}\right) + \beta_L\left(e_{Lk}/e_{L,avg}\right)\right] \tag{10}$$

where

A_k = the relative attraction of a destination;
φ = a calibration constant;
e_{sk}, e_{Hk}, e_{Lk} = the employees of a destination city in the service, heavy, and light industries respectively;
$e_{s,avg}$, $e_{H,avg}$, $e_{L,avg}$ = the number of employees in the service, heavy, and light industries respectively in the average city of the network; and
β_S, β_H, β_L = the trip attraction characteristics of each employment type (these were found to be 0.452, 0.363, and 0.185 respectively from Air Canada data).

Systems Graph

The systems graph is a set of terminal graphs connected at the vertices to form a one-to-one correspondence with the components of a physical system. Figure 4 shows three alternate Toronto airport locations. Figure 5 shows the systems graph for the expansion of the existing terminal. The elemental numbers are given in Table 1.

Systems Equations

To construct the travel demand model, it is necessary to derive both the chord and branch formulation equations of the system. For purposes of the study, it was assumed that trips outside the Toronto-Montreal-Ottawa triangle would remain unaffected by the airport location. The three origins used in the example were the area northwest of Toronto, the Hamilton-Niagara Peninsula, and Metropolitan Toronto (Fig. 4).

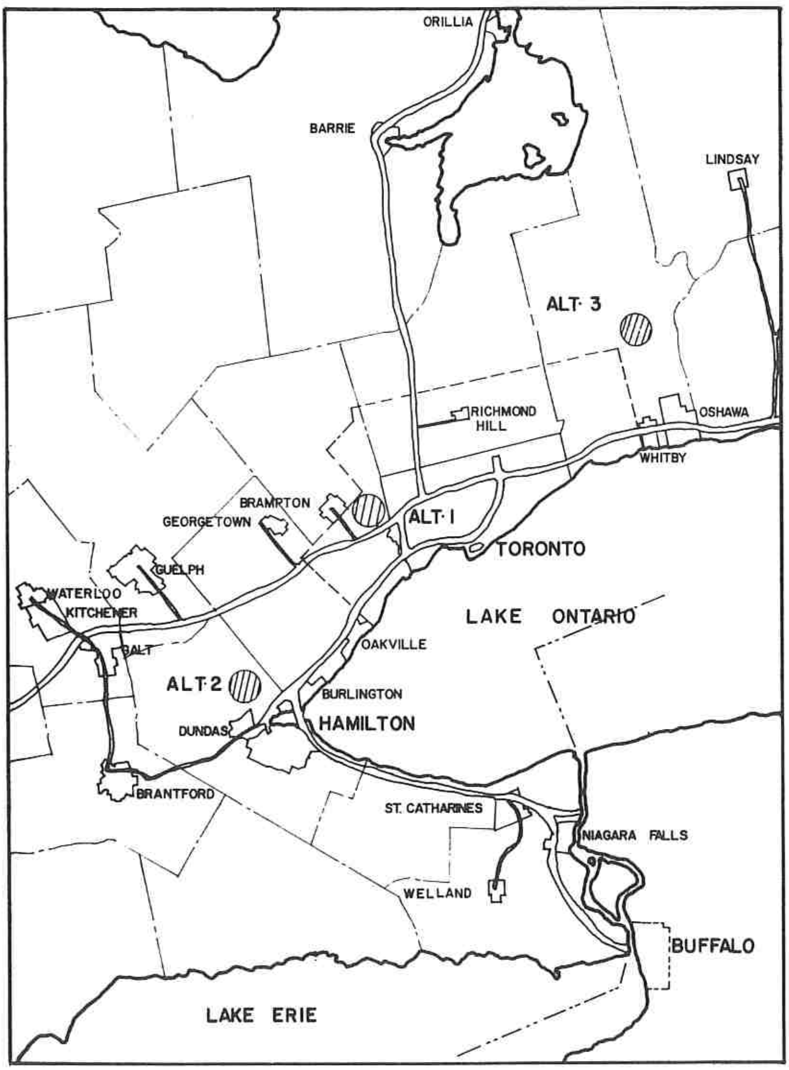


Figure 4. Alternatives for the Toronto International Airport system.

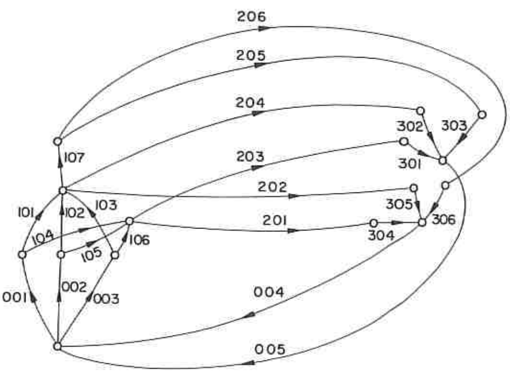


Figure 5. System graph for airport location alternative.

The resistance and attraction values are listed in Table 1. The airport access costs and times were as stated previously. The chord formulation model is of the form

$$[A]\,[R]\,[A^T]\begin{bmatrix} Y_c \\ Y_{c1} \end{bmatrix} + \begin{bmatrix} O \\ U \end{bmatrix} X_{c1} = 0 \quad (11)$$

These reduce to

$$[Ƶ]\begin{bmatrix} Y_c \\ Y_{c1} \end{bmatrix} = 0$$

where

$Ƶ$ = the coefficients of the matrix triple product;
Y_c = the unknown flow values;
Y_{c1} = the known flow values for the three origin areas; and
X_{c1} = the pressure or travel potential for the three origins.

The unknown flow values are calculated. These values are substituted in the last three equations of the set (18), and the travel potentials of the origin area are derived. The flow values for alternative 1 were verified with actual data. These are given in Table 2.

New resistance values for alternatives 2 and 3 then are employed. With these new resistance values and the derived travel potentials, the branch formulation models are solved to determine the traffic consequences associated with each airport location. The branch formulation equations are of the form

$$\begin{bmatrix} U \\ O \end{bmatrix} Y_{c1} + [B]\,[R]\,[B^T]\begin{bmatrix} X_{B1} \\ X_B \end{bmatrix} = 0 \quad (12)$$

TABLE 1

VALUES FOR LINK RESISTANCES, ALTERNATIVE 1

Link No.	Type	Resistance	Attraction	Flow
001	Toronto North-West			50,000
002	Hamilton-Niagara			61,400
003	Toronto (Metro)			223,000
004	Ottawa		1.253	
005	Montreal		4.038	
101	Air-highway access	2,540		
102	Air-highway access	2,730		
103	Air-highway access	1,540		
104	Rail access	4,565		
105	Rail access	4,940		
106	Rail access	1,310		
107	Airport process	760		
201	Rail link	10,715		
202	Highway link	4,565		
203	Rail link	19,450		
204	Highway link	3,460		
205	Air link	3,680		
206	Air link	3,340		
301q	Rail egress	5,850		
302	Highway egress	5,900		
303	Air egress	1,160		
304	Rail egress	5,850		
305	Highway egress	5,850		
306	Air egress	1,100		

Sources: 1. Regional Studies, Department of Highways of Ontario.
2. Airline Statistics 1964, Air Transport Board.
3. Point-to-Point Passenger Volumes, Canadian National Railways.

TABLE 2
DERIVED TRAFFIC VOLUMES FOR ALTERNATE AIRPORT SITES

Origin Toronto	One-Way Business Trips (persons/yr)					
	Montreal			Ottawa		
	Air	Rail	Road	Air	Rail	Road
Alternate 1	97,900	70,439	62,602	52,310	18,042	40,427
Alternate 2	103,300	69,094	60,064	59,500	16,100	38,872
Alternate 3	89,300	72,094	64,064	43,500	20,082	42,383

Note: Base year 1964.

where X_B is the known travel potentials for the three origins.

Having solved for X_{B1} (the unknown pressures), the link volumes are derived from the terminal equations. The results, in terms of 1964 business trips, are given in Table 2.

Model Results

The results of the branch model are as would be anticipated. The lower access associated with alternative 2 (Toronto Airport plus Hamilton Airport) resulted in a generated air traffic volume of about 1,500 yearly business passengers on the Toronto-to-Montreal air link as compared to alternative 1. [Generated traffic equals new air volume minus old air volume minus diverted traffic; i.e., 103,300 - 97,900 - (70,439 - 69,094) + (62,602 - 60,064) = 1,497.] Furthermore, 1,345 annual trips were diverted from rail and 2,558 from automobile. For the Toronto-Ottawa city pair, 3,693 new trips were generated on the air mode, whereas 1,942 were diverted from rail and 1,555 were diverted from automobile.

Under alternative 3 (new regional airport east of Toronto), it is anticipated that about 5,500 business trips per year would not be made from Toronto to Montreal as compared to alternative 1 [total air traffic lost minus traffic diverted to auto and rail; i.e., (97,900 - 89,300) - (72,094 - 70,439) - (64,064 - 62,602) = 5,483]. On the Toronto-Ottawa route, about 4,800 annual business trips would not be made, whereas 2,040 trips would be lost to rail and 1,956 trips would be lost to automobile.

The accuracy of these results, of course, cannot be verified. The travel elasticities, however, appear reasonable.

APPLICATION TO AIRPORT LOCATION PROCESS

The user travel benefits for the airport location example were calculated. The calculation was based on the assumption that the travel link resistance measures represent the perceived cost of travel for the aggregate of travelers. The measurement of benefits was for the Toronto-to-Montreal route only.

TABLE 3
PERCEIVED COSTS AND TOTAL CORRIDOR TRIPS FOR AIRPORT ALTERNATIVES
(Toronto to Montreal only)

Alternative	Total Perceived Costs Per Trip[a] (dollars)				Total Annual Business Trips[b] (000's of person trips)			
	Air[c]	Rail[c]	Road[c]	Total[c]	Air	Rail	Road	Total
1	27.60	51.80	29.40	108.80	97.9	70.4	62.6	230.9
2	27.00	50.70	28.10	105.80	103.3	69.1	60.1	232.5
3	28.10	54.60	30.90	113.60	89.3	72.1	64.1	225.5

[a]Perceived costs are total of access, link, and egress resistance values. Perceived costs by aggregate of travelers.
[b]Business trips are originating in Toronto area for Montreal only.
[c]Total perceived costs are weighted by number of normalized trips originating in Toronto for Montreal. The procedure is outlined in Equation 4.

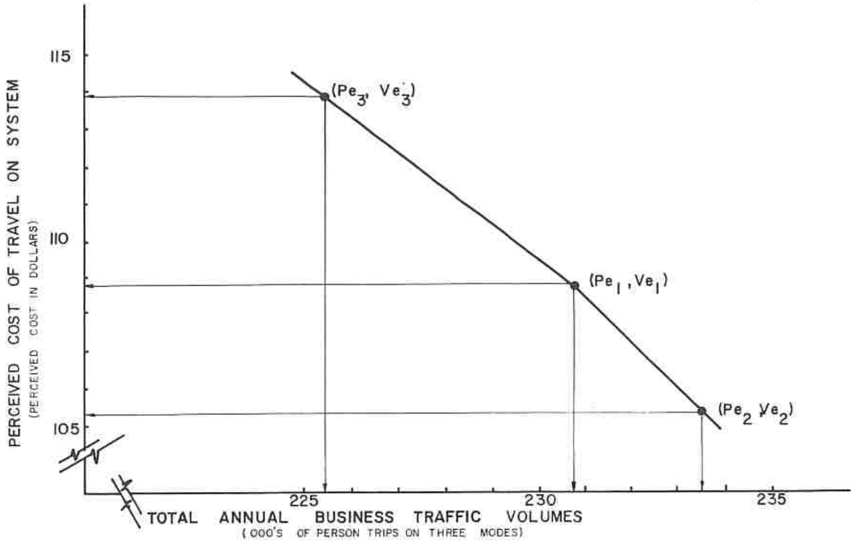


Figure 6. Equilibrium demand points for alternative Toronto airport locations.

The perceived costs and total corridor business trips are given in Table 3. The entries of perceived costs are the weighted resistances of the access, travel route, and egress links. The weighting was achieved by normalizing the total trips on each mode. The total perceived costs and the total business trips associated with each alternative are the derived equilibrium demand points for that alternative.

Figure 6 shows the equilibrium demand points for each alternative. If alternative 1 (Toronto International Airport as proposed) is taken as the do-nothing alternative, then the net user benefits of alternative 2 (Toronto plus Hamilton) and alternative 3 (regional airport east of Toronto) can be determined.

The method of benefit calculation is the Hewes-Oglesby method (19). Comparing alternative 3 with alternative 1, a net disbenefit would result. It would be equal to

$$\tfrac{1}{2}\left(P_{e1} - P_{e3}\right)\left(V_{e3} + V_{e1}\right) = -1{,}095{,}360.00$$

The value is a perceived dollar value for the year 1964. Undoubtedly, all other things being equal, this proposal should be discarded.

The comparison of alternative 2 with alternative 1 produces a net benefit equal to

$$\tfrac{1}{2}\left(P_{e1} - P_{e2}\right)\left(V_{e2} + V_{e1}\right) = 694{,}500$$

The positive perceived value is for the one year. To determine the total user benefits that accrue over the study period, yearly passenger forecasts for pleasure as well as business travel must be made, and the associated benefits calculated. These values must be reduced to present-day terms. The planner then must decide if the additional investment for a Hamilton airport is justified in light of the user benefits (assuming all other things are equal).

SUMMARY

A systems model to forecast air traffic demands was described. One of its uses, within the total airport location study framework, is to assess the impact of airport location on the total traffic potential for short-haul air trips. An example of the method applied in the Toronto region indicates that the procedure can be used effectively.

ACKNOWLEDGMENTS

The research presented in this paper has been sponsored by The National Research Council of Canada and The Transportation Policy and Research Branch, Canada Department of Transport.

REFERENCES

1. Toronto International Airport Master Plan. Prepared for the Canada Department of Transport by John B. Parkin Associates Ltd., Waterloo, 1967.
2. Montreal International Airport Master Plan. Prepared for the Canada Department of Transport by Kales, Peat, and Marwick, Montreal, 1967.
3. Toronto International Airport System. Prepared for the Canada Department of Transport by P. M. Pearson, Waterloo, 1968.
4. Hutchinson, B. G., and Pearson, P. M. An Evaluation of Ground Transportation Requirements for Major Airport. Canadian Transportation Research Forum, 1968.
5. Pearson, P. M., and Hutchinson, B. G. A Location Analysis of the Toronto and Region Airport System to 1968. Prepared for John B. Parkin Associates Ltd., Waterloo, 1968.
6. Effective Community Air Travel Potential. Civil Aeronautics Administration, Washington, D. C., 1950.
7. Mylroie, Willa. Evaluation of Intercity Travel Desire. HRB Bull. 119, 1956, pp. 69-94.
8. Tanner, J. C. Factors Affecting the Amount of Travel. Road Research Technical Paper 51, Road Research Laboratory, London, 1961.
9. Demand for Intercity Passenger Travel in the Washington-Boston Corridor. Systems Analysis and Research Corporation, Boston, 1965.
10. Blackburn, A. J. Studies in Travel Demand. Series of lectures delivered at Univ. of Waterloo, 1969.
11. Wohl, M., and Martin, B. Traffic System Analysis for Engineers and Planners. McGraw-Hill, New York, 1967.
12. Ellis, J. B. The Description and Analysis of Socio-Economic Systems by Physical Systems Techniques. Michigan State Univ., PhD dissertation, 1965.
13. Dodson, R. M. A Systems Theory Model of Population Migration Streams. Univ. of Waterloo, MS thesis, 1967.
14. Pearson, P. M. The Planning and Evaluation of Intercity Travel Systems. Univ. of Waterloo, PhD dissertation, 1969.
15. The Vancouver Study. Air Canada, Montreal, 1966.
16. Seshu, S., and Reed, M. B. Linear Graphs and Electrical Networks. Addison-Wesley, Reading, Mass., 1961.
17. Roe, P. H. O'N. Networks and Systems. Addison-Wesley, Reading, Mass., 1966.
18. Model Airport Zoning Ordinance. U. S. Federal Aviation Agency, Washington, D. C., 1961.
19. Oglesby, C. H., and Hewes, L. I. Highway Engineering. 2nd Ed. John Wiley and Sons, New York, 1963.

Transportation: The Link Between People and Jobs

HAROLD KASSOFF, Bureau of Public Roads, Federal Highway Administration, U.S. Department of Transportation; and
HAROLD D. DEUTSCHMAN, Newark College of Engineering

•THE CONCENTRATION of low-income households in the core areas of the nation's cities, coupled with a growing trend towards dispersion of employment opportunities, particularly in the unskilled and semiskilled categories, is resulting in a growing spatial mismatch of low-income residential areas and the location of available jobs. The problem is compounded by the general reliance of poverty-level households on public transportation systems that typically do not provide adequate access to outlying suburban areas. Although it is certainly clear that the improvement of living standards for more than 10 percent of the nation's population demands much more than improvements in accessibility to employment, evidence exists in some areas that the inability of workers to reach jobs as a result of transportation constraints can be a major factor in limiting their economic status. It is the purpose of this study to provide some insight and dimension to the nature of this people-job-transportation relationship in terms of its implications for poverty-level families.

The study was conducted for the New York metropolitan area using data available from the 1963 Tri-State Transportation Commission Home Interview Survey. The survey consisted of a 1 percent sample of households drawn from the 22 counties lying in those portions of New York, New Jersey, and Connecticut that comprise the Tri-State Study Area (Fig. 1). The data compiled reflect extensive socioeconomic as well as travel information for the residents of the study area.

For the purposes of this project, only those permanent-resident households with the head in the labor force were selected from the Home Interview file. This included employed heads of households as well as those who were unemployed but seeking employment when the survey was conducted. Not included were those heads of households classified as retired, students, or housewives; such persons are generally not affected significantly by the relative availability of employment opportunities. It must be emphasized that the labor force and employment statistics quoted in this report apply only to heads of households.

The research was structured into three major phases. The first was aimed at providing a descriptive profile of the social and travel characteristics of low-income households and an indication of the nature of the variation of some of these characteristics with household income. Such factors as age of head of household, family size, trip rates, occupation, industry, auto ownership, and residential and employment mobility are examined in terms of the financial status of the household.

The second phase of the study is concerned with the spatial distribution of low-income households in terms of places of residence and places of employment. As part of this analysis, graphic displays were prepared that show the concentrations of low-income homesites and worksites in the study area on a square-mile basis. Also shown is the geographical distribution of low-income unemployed heads of households. Statistics were compiled that reflect the percent distribution of households by income within counties both for places of residence and for employment. The availability of a private automobile was considered an important indicator of household travel potential and therefore was introduced as a classifying variable in some of the tabulations.

Paper sponsored by Committee on Origin and Destination.

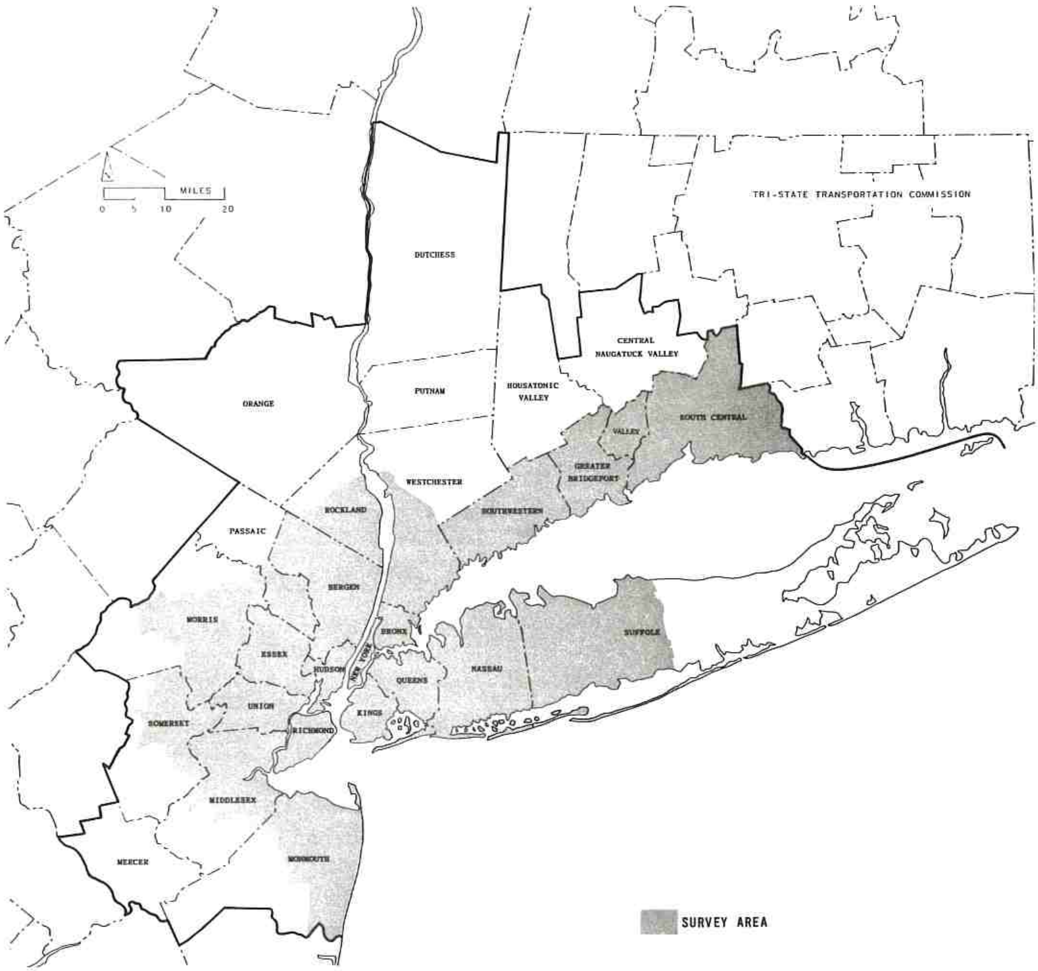


Figure 1. Tri-State Home Interview Survey area.

The final phase of the study consists of a profile of the journey-to-work characteristics of households according to income classification. Included are an analysis of trip length in terms of time as well as distance, percent transit usage, reverse commuting by residents of New York City, and percent of households living and working in the same county. Once again auto availability was used in many instances to further stratify households into groups of unique characteristics.

SOCIOECONOMIC PROFILE

There were approximately 16 million persons living in the 3,600-square-mile study area when the survey was conducted in 1963. About 4.25 million heads of households were in the labor force, 4 million of which were employed within the study area (Tables 1 and 2). A little more than 10 percent of these heads of households reported annual family incomes of under $4,000 per year. In New York City this figure increased to about one out of every five, or 20 percent. Approximately three-quarters of the study area's low-income head-of-household labor force resided within New York City, whereas somewhat less than that fraction of total low-income employment was located there.

TABLE 1

PERCENT DISTRIBUTION OF LABOR FORCE BY HOUSEHOLD INCOME

Residential Location	Household Income Percentages			Total Labor Force (thousands)
	Under $4,000	$4,000-$10,000	Over $10,000	
New York City (excl. Richmond)	18.4	61.7	19.9	2,111
Outside New York City (incl. Richmond)	6.7	62.5	30.8	2,132
Study area	12.5	62.0	25.5	4,243

Note: Data include heads of households only, both employed and unemployed.

TABLE 2

PERCENT DISTRIBUTION OF EMPLOYMENT BY HOUSEHOLD INCOME

Employment Location	Household Income Percentages			Total Employment[a] (thousands)
	Under $4,000	$4,000-$10,000	Over $10,000	
New York City (excl. Richmond)	13.4	60.2	26.4	2,328
Outside New York City (incl. Richmond)	7.1	67.1	25.8	1,715
Study area	10.7	63.1	26.2	4,043

[a]Includes heads of households only.

The Tri-State study area contains a number of major urban centers besides New York City, including Newark, Jersey City, New Haven, and Bridgeport. In view of the relative dominance of New York City within the metropolitan region, however, most of the geographic stratification in this report is by location either within or outside New York. For analysis purposes the statistics in this report referring to New York City include only the four boroughs of Manhattan (New York County), Brooklyn (Kings County), Queens, and the Bronx. Richmond, because of its low density pattern of development, was considered as part of the rest of the study area outside of New York City. (The population of Richmond is only 3 percent of the total population of the city.) A number of tabulations were produced on a county basis (Tables 12-16) and these have been included in the Appendix.

Employment Characteristics

For low-income households with the head in the labor force, the key to improvement of economic well-being is the expansion of employment opportunities. In this respect, occupation becomes a key factor. It is not surprising to find that over 75 percent of low-income heads of households are either blue-collar workers or are unemployed, compared with about 55 percent of middle-income and 25 percent of high-income workers (Table 3). What is perhaps most startling is that almost one in every five heads of households in the low-income category (100,000 persons) is unemployed and actively seeking a job, whereas only slight fractions of middle- and high-income household heads are unemployed. (The definition of unemployed as used in the Home Interview Survey referred to individuals who were without jobs but actively seeking employment.) This situation arises in part as a consequence of shifting patterns in employment opportunities. The trend towards automation in many industries is resulting in both unemployment and a surplus of jobs. The problem in large part is that the talents of the unemployed do not generally match the requirements of the available

TABLE 3

OCCUPATIONAL DISTRIBUTION OF HEADS OF HOUSEHOLDS BY HOUSEHOLD INCOME

Household Income	Percent Distribution								Total Labor Force
	Managers	Clerical	Sales	Crafts-men	Opera-tives	Ser-vice	Labor-ers	Unem-ployed	
Under $4,000	11.4	11.1	3.1	9.0	21.5	16.5	4.7	18.3	532,000
$4,000 to $10,000	24.5	13.0	6.1	21.5	18.9	9.2	4.2	2.2	2,634,000
Over $10,000	60.5	5.6	10.0	12.1	6.7	3.3	1.2	0.6	1,078,000
All income classes	32.1	10.9	6.7	17.6	16.1	8.5	3.5	3.8	4,244,000

TABLE 4

INDUSTRY DISTRIBUTION OF HEADS OF HOUSEHOLDS BY HOUSEHOLD INCOME

Household Income	Percent Distribution								Total Employed
	Con-struction	Manufac-turing	Utilities, Commun., Transp.	Whole-sale	Retail	Finance Insur., Real Estate	Prof. and Service	Public Admin.	
Under $4,000	3.4	30.2	4.7	2.6	15.2	7.8	31.9	4.2	434,000
$4,000 to $10,000	7.1	31.5	11.7	4.7	13.4	6.0	17.7	7.9	1,552,000
Over $10,000	6.3	30.4	8.7	6.2	10.5	8.9	23.4	5.6	1,057,000
All income classes	6.5	31.1	10.2	4.8	12.8	6.9	20.7	7.0	4,043,000

jobs. The sharp differences between the occupational distribution of heads of low-income households and the rest of the labor force, coupled with the inordinately high unemployment rate in the former group, amplify the point.

A distribution of employment by industry type, for each income class, is given in Table 4. The manufacturing industry stands out as a major employer in all income categories. For low-income heads of households, the service industry accounts for a high percentage of employment compared with other income groups. Surprisingly, though, the relative percentages of heads of households employed in most of the industries do not vary sharply with income.

Family Characteristics

Household Size—Though the association of poverty with households is frequently not considered in terms of family size, it is clear that the number of persons in the family unit has a great bearing upon the standard of living attainable from a given income. In general, there seems to be a positive correlation between family size and income (Fig. 2). The degree of association varies somewhat by location; households residing outside New York City are generally larger than those within the city for a given income class. The tendency towards declining family size with decreasing income is preserved when households are stratified by age of head, as given in Table 5. For the under 35 age group, 31 percent of the low-income households consist of only one person whereas for middle- and high-income households the percents are only 9 and 4 respectively. Similar proportions apply to the over 35 age group. Thus, it appears that in the study area about one-third of the households with incomes under $4,000 per year are single-person households for which the classification of "poverty" may be subject to some question.

Age of Head of Household—The relationship between income and age of head of household was examined by looking at the distribution of households by income class within each age group (Fig. 3). Although it is apparent that low-income households

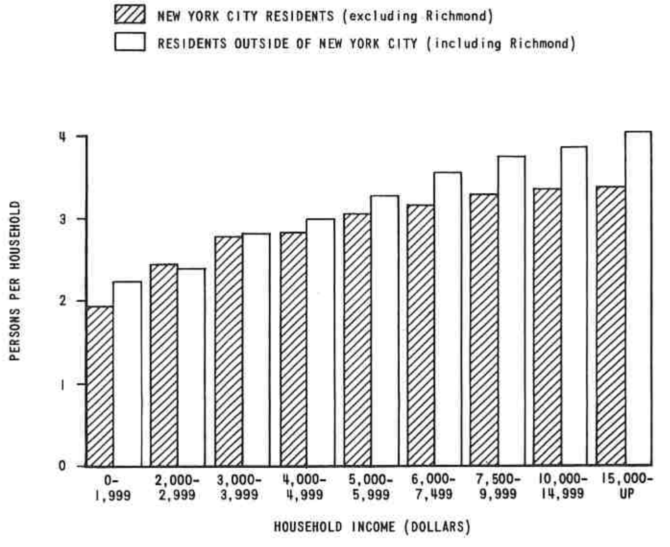


Figure 2. Persons per household by income and place of residence.

TABLE 5

PERCENT DISTRIBUTION OF PERSONS PER HOUSEHOLD STRATIFIED BY INCOME AND AGE GROUP

Income	Percent Distribution of Persons per Household							
	Under 35 Age Group				Over 35 Age Group			
	1	2	3	4+	1	2	3	4+
Under $4,000	31.3	17.1	18.6	33.0	37.2	28.9	12.4	21.5
$4,000 to $10,000	8.9	18.4	23.7	49.0	10.0	28.0	19.2	42.8
Over $10,000	3.8	31.0	17.6	47.6	3.0	24.4	22.2	50.4
All income classes	11.9	20.1	21.9	46.1	10.9	27.0	19.4	42.7

TABLE 6

PERCENT DISTRIBUTION OF AGE OF HEAD OF HOUSEHOLD[a] BY HOUSEHOLD INCOME

Income	Age Group						Total Households[a]
	Under 25	25-34	35-44	45-54	55-64	Over 65	
Under $4,000	12.7	21.1	21.4	20.1	17.0	7.7	531,737
$4,000 to $10,000	4.1	22.3	29.1	25.2	15.9	3.4	2,634,607
Over $10,000	1.1	13.3	28.0	33.3	20.2	4.1	1,077,577
All income classes	4.4	19.8	27.9	26.6	17.2	4.1	4,243,921

[a]Includes members of labor force only.

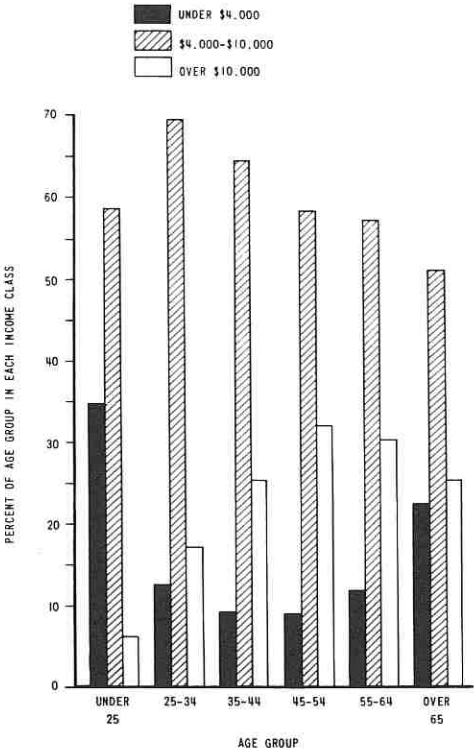


Figure 3. Percent distribution of household income by age of head of household.

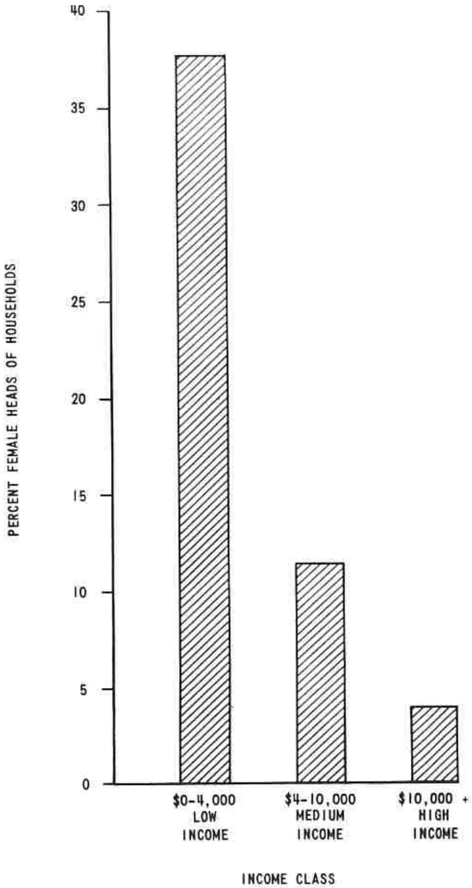


Figure 4. Percent of female heads of households in the labor force by income.

are in the minority in virtually every age group, the most interesting feature of the distributions are their relative shapes. The middle- and high-income households peak in terms of their relative presence in each age group in the 25 to 34 and 45 to 54 categories respectively, both trailing off somewhat in the youngest and oldest age categories. Low-income households are distributed quite differently, with their relative presence within each age group clearly peaking in the very young and very old age groups. In terms of percentage figures, low-income households comprise 35 percent of the under 25 group and almost 25 percent of the over 65 age group (note that retired heads of household are not included), whereas in the 35 to 44 and 45 to 54 age categories less than 10 percent are low-income households. Further analysis revealed that 20 percent of all low-income households fell into these two extreme age groups, whereas for middle- and high-income households these figures were 8 and 5 percent (Table 6).

It seems clear that compared to the rest of the population, a disproportionately large share of low-income households are clustered in the very young and very old age groups. This is not to minimize, however, the significance of the 80 percent of the low-income group that remains in the age 25 to 64 categories. The distribution does indicate, however, that a significant share of heads of households with annual incomes under $4,000 are either near retirement or have just entered the labor force and have not developed to their full earning potential.

Sex of Head of Household—The distribution of sex of head of household by income class provides an interesting and significant insight into the composition of low-income

households. As shown in Figure 4, nearly 40 percent of the heads of low-income households in the labor force are females. This compares with 11 percent of middle-income and 4 percent of high-income heads of households who are female.

An analysis of unemployed low-income heads of households showed that 35 percent of those with incomes under $4,000 per year who were out of work at the time of the survey were females. The unemployment rate among female low-income heads of households was virtually the same as for males, the former being 17 percent and the latter, 19 percent.

The high proportion of female heads of households in the low-income group is a consequence of social conditions that will not be discussed here. What is of major importance is how this relates to the notion of improving job opportunities for the poor. For example, a public works program that intended to increase the supply of jobs in the manufacturing and construction industries would have little effect upon four-tenths of the heads of households in the low-income labor force. There are transportation implications as well. Females are usually more reliant on transit and less willing or able to travel long distances to work (1). Such factors must be considered carefully in the development of programs to provide jobs or transportation to jobs for low-income households.

Residence and Employment Changes (1960-1963)

In analyzing the relationship between people, jobs, and transportation, it is useful to examine the dynamics of residential and employment mobility. The willingness of low-income households to change residence and to leave current low-paying jobs are factors to be considered in attempting to improve their accessibility to job opportunities. In this section, the frequency of residential and employment changes that occurred between the 1960 census and the 1963 Home Interview Survey is examined with respect to both income and age of head of household.

Residential Mobility—The relative residential mobility of households, stratified by income class as well as age of head, is given in Table 7 in terms of moves per thou-

TABLE 7

NUMBER OF RESIDENCE CHANGES (1960-63) PER THOUSAND HOUSEHOLDS BY INCOME AND AGE GROUP

Age of Head of Household	Under $2,000	$2,000 to $2,999	$3,000 to $3,999	$4,000 to $4,999	$5,000 to $5,999	$6,000 to $7,499	$7,500 to $9,999	$10,000 to $14,999	$15,000 and Over	All Income Classes
Under 25	1,261	1,518	1,127	1,276	1,154	1,264	1,281	1,517	1,562	1,269
25-34	1,140	1,013	807	744	749	772	782	848	832	798
35-44	924	635	497	426	358	333	312	360	413	375
45-55	476	434	394	327	237	210	202	207	258	245
55 and over	236	229	187	186	170	142	138	139	295	178
All age groups	758	665	536	479	424	408	367	360	363	421

TABLE 8

NUMBER OF JOB CHANGES (1960-63) PER THOUSAND HOUSEHOLDS BY INCOME AND AGE GROUP

Age of Head of Household	Under $2,000	$2,000 to $2,999	$3,000 to $3,999	$4,000 to $4,999	$5,000 to $5,999	$6,000 to $7,499	$7,500 to $9,999	$10,000 to $14,999	$15,000 and Over	All Income Classes
Under 25	270	741	498	615	622	686	603	539	688	572
25-34	403	463	433	406	414	434	437	464	430	432
35-44	424	365	246	252	253	228	241	298	291	261
45-54	271	236	213	160	149	156	172	192	176	175
55 and over	202	179	184	126	145	115	116	155	220	156
All age groups	292	353	293	262	263	263	247	264	243	263

sand households. The analysis indicates that frequency of moves tends to decline with increasing income in the older age groups. In the younger age groups, the middle-income households appear to be the most stable, whereas the low- and high-income households have a greater frequency of residential moves. Not surprisingly, the rate of residence changes declines with increasing age regardless of income group. In general, the analysis shows that low-income households have a greater tendency to change residence than the rest of the population.

Employment Changes—The pattern of employment changes per thousand households as a function of income and age of head of household is somewhat less distinct than the distribution of rates of residential changes. There seems to be no systematic variation in job changes by the head of household with household income. As was the case with shifts in residence, however, the rate of employment changes declines with increasing age in virtually all income classes (Table 8).

Travel Characteristics

Auto Availability—The availability of a private automobile is a key determinant of a household's travel behavior and is strongly influenced by income. Figure 5 shows the decrease in percent of households with no autos available as household income increases. Of New York City low-income households, less than one in five has a car available. This is contrasted with the highest income class residing outside New York City of which almost 99 percent own at least one automobile. The significantly higher percent of zero-auto households for New York City residents compared with the rest of the population, regardless of income class, is a reflection of both New York City's extensive transit system and the relative expense and inconvenience of maintaining a private auto within the city. The graph clearly indicates that reliance on modes of travel other than the automobile is highest among the lowest income groups within New York City.

A further analysis of auto availability within the low-income group by employment status shows that the percent of households with zero autos is significantly higher when

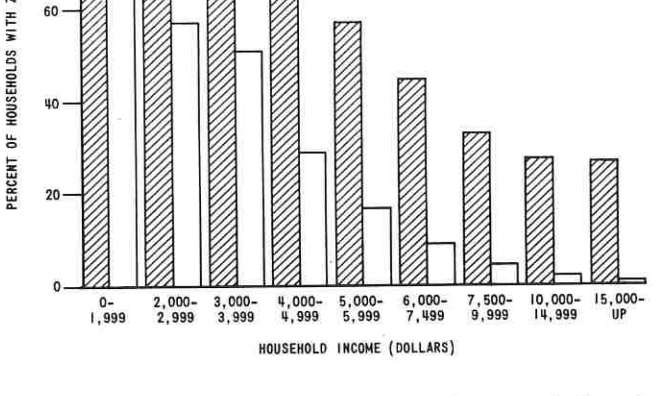


Figure 5. Percent of zero-auto households by income and place of residence.

the head is unemployed than when he holds a job (Table 9). The need for an automobile in areas of generally poor transit service is demonstrated by the fact that as much as 53 percent of the employed and 33 percent of the unemployed low-income households residing outside New York City have at least one private auto available.

TABLE 9

PERCENT AUTO AVAILABILITY FOR LOW-INCOME HOUSEHOLDS BY EMPLOYMENT STATUS

Location	Employment Status	Autos Available	
		0	1+
New York City (excl. Richmond)	Employed	85.8	14.2
	Unemployed	92.4	7.6
Outside New York City (incl. Richmond)	Employed	53.3	46.7
	Unemployed	67.0	33.0

Number of Trips—The relationship between household travel, in terms of number of trips and household income is shown in Figure 6. The graph demonstrates that households with higher incomes make more trips than those in the lower income groups, and that the trip rates for households residing outside New York City are progressively higher than for New York City residents as income increases. (Walking trips are not included.)

To account for the correlation between household size and household income, trips per person also were analyzed in relation to household income. The results shown in Figure 7 indicate that the positive relationship between travel and income is preserved, even on a per-person basis. The rate varies from a low of about one trip per person for households earning less than $2,000 per year to a high of three trips in the $15,000-and-over income category.

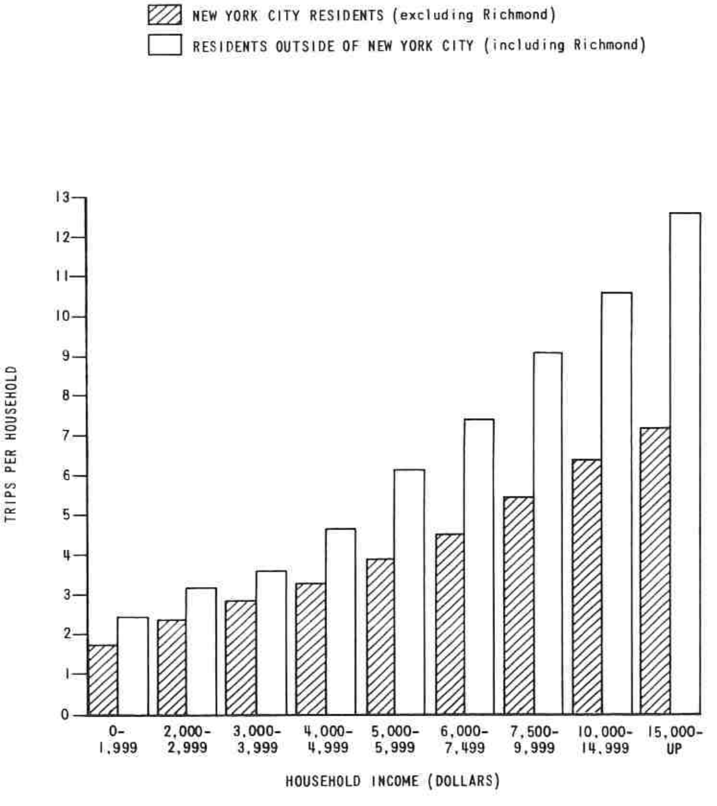


Figure 6. Trips per household by income and place of residence.

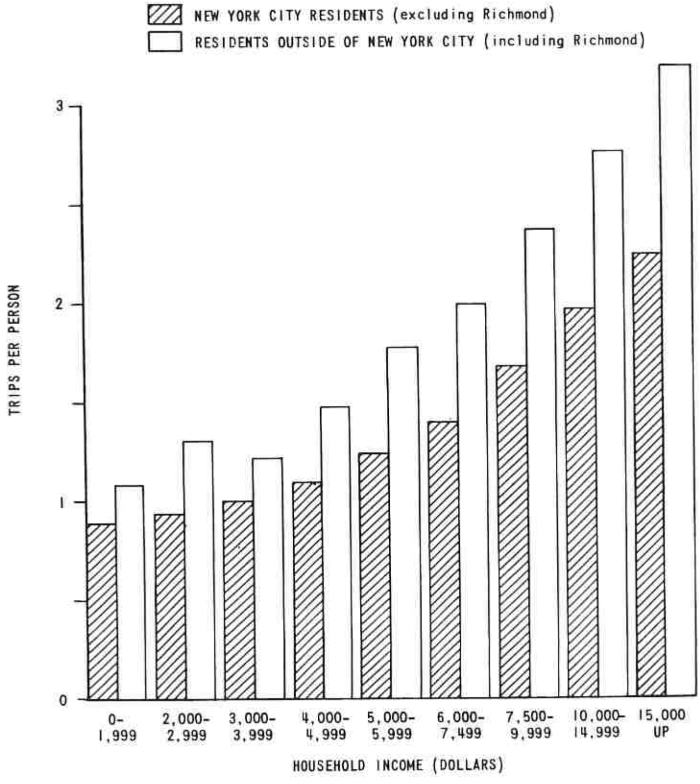


Figure 7. Trips per person by income and place of residence.

Thus, the travel mobility of low-income households, measured in terms of number of trips, is somewhat restricted in comparison with the wealthier segments of the population. Although travel per se is merely a means to an end, the implications are that the opportunity to engage in activities is more constrained for low-income households.

SPATIAL DISTRIBUTION

The spatial separation of place of residence from place of employment brings about the need for transportation to and from work. Although logic may dictate that this separation should be minimized, the pattern of growth in population and employment in most urban areas seems to defy this seemingly sound conclusion. In most major metropolitan areas the geographical distribution of middle- and upper-income households has been shifting to the suburbs while the populations of core areas increasingly are dominated by households with relatively low incomes. On the other hand, there is a trend toward a dispersion of major manufacturing employers who provide a large share of job opportunities for unskilled or semiskilled workers; whereas high-income, white-collar jobs tend to remain within centrally located areas.

This section examines the spatial distribution of homesites and worksites of the low-income labor force as it existed in the Tri-State area in 1963. Even though this analysis represents a single cross section in time, it should be considered in the context of the shifting pattern of urban development just described.

Distribution of Low-Income Homesites

As shown by the map in Figure 8, the major concentration of members of the low-income labor force is in the core area of the region. In New York City almost one in five heads of households in the labor force is in the low-income category (in Manhattan the ratio is more than one in four) whereas in the rest of the study area considerably less than one in 10 fall into this group (Table 12, Appendix). Outside of New York City, there is a general tendency towards higher percentages of low-income households in the more densely developed counties.

It is difficult to foresee any drastic shifts in the distribution of low-income households within the near future. Despite such efforts as open-housing legislation and rent subsidy programs, it is likely that the poor will remain in core areas where housing costs are relatively low, where the ownership of a private vehicle is not a necessity, where public welfare services are readily available, and where ethnic and economic barriers are not prohibitive.

Distribution of Low-Income Worksites

As is the case with the distribution of low-income homesites, the major areas of low-income employment of heads of households center around New York City (Fig. 9). Although 400,000 low-income heads of households reside within the city, only a little

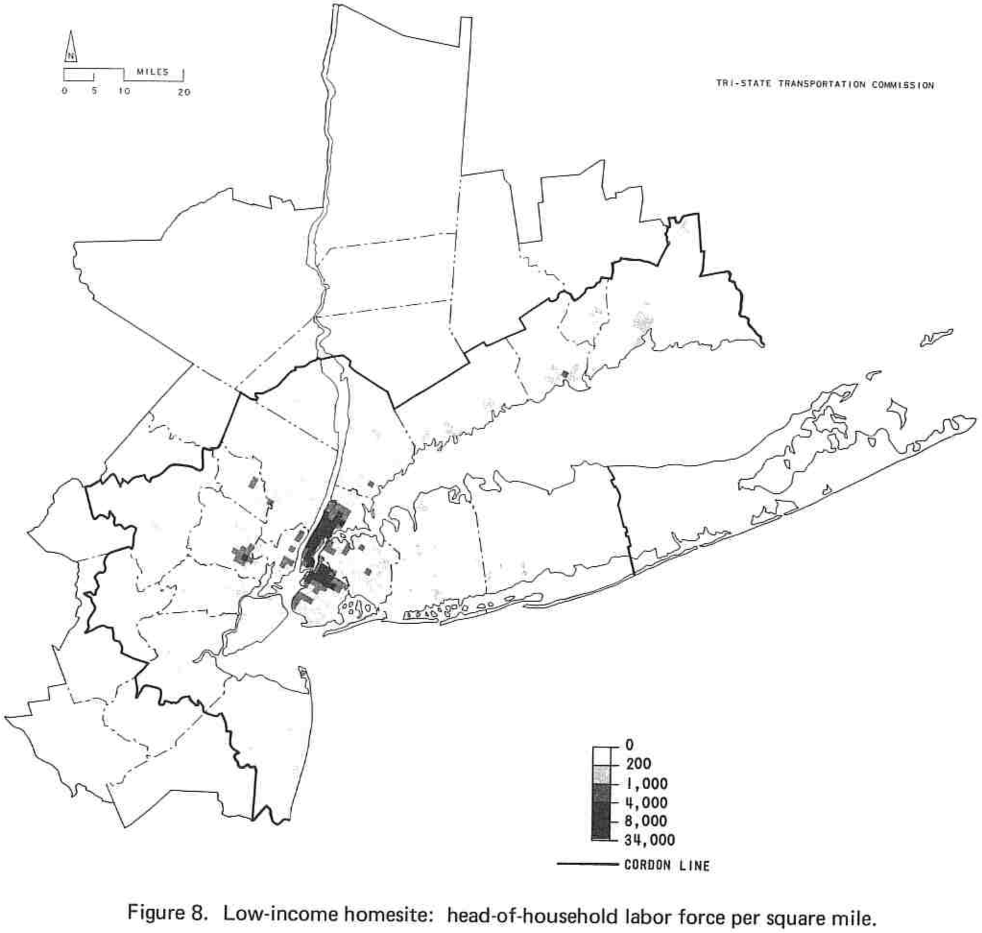


Figure 8. Low-income homesite: head-of-household labor force per square mile.

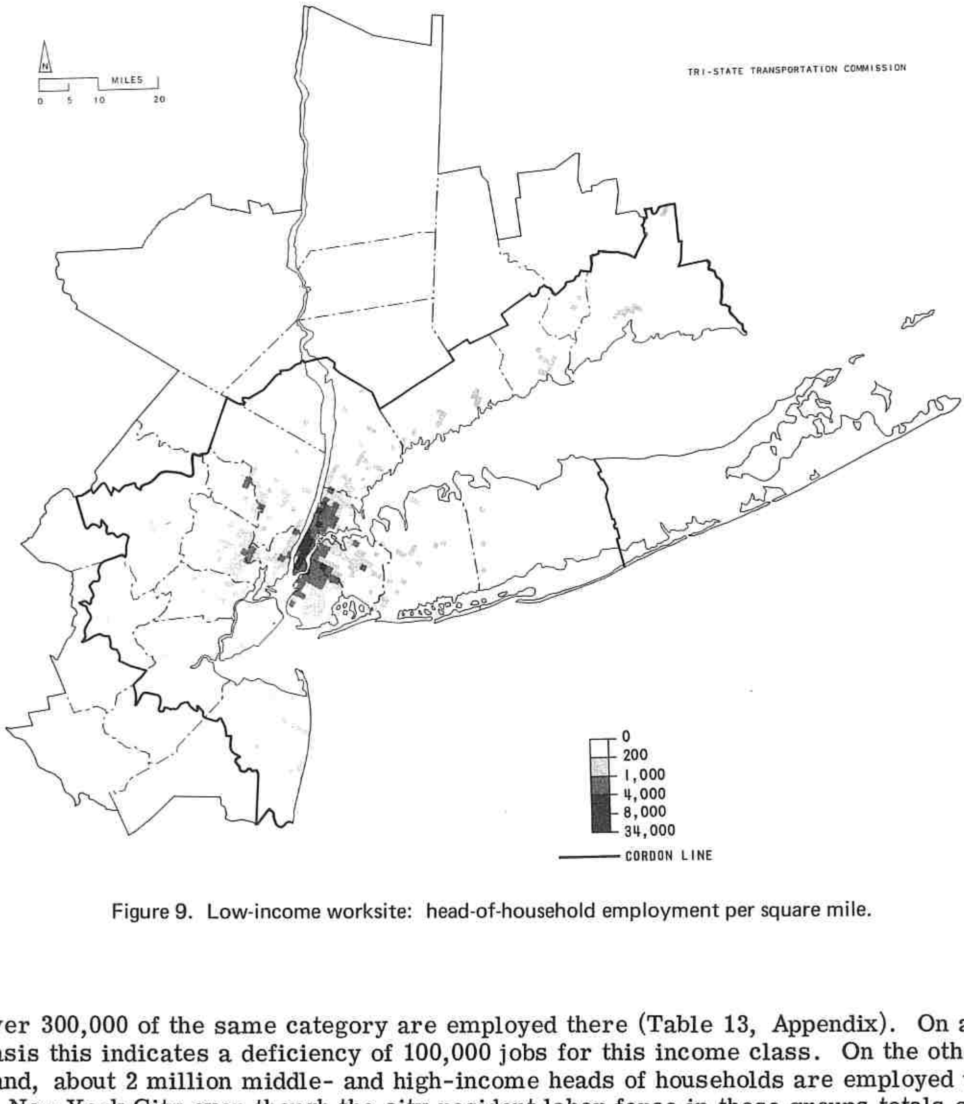


Figure 9. Low-income worksite: head-of-household employment per square mile.

over 300,000 of the same category are employed there (Table 13, Appendix). On a net basis this indicates a deficiency of 100,000 jobs for this income class. On the other hand, about 2 million middle- and high-income heads of households are employed within New York City even though the city resident labor force in these groups totals only about 1.7 million. It can be concluded, therefore, that on a net basis New York City imports workers (who are heads of households) to fill jobs in the upper-income categories while a significant portion of the low-income labor force residing within the city remain unemployed or seek positions elsewhere. This conclusion emphasizes the importance of accessibility to job opportunities for low-income households, many of whom must reverse commute from New York City.

Distribution of Low-Income Unemployed

Low-income unemployed heads of households represent only about 2 percent of the total labor force. In New York City this figure rises to over 3 percent whereas outside the city only slightly over 1 percent are in this category. Considering the absolute size of the study area labor force, however, these persons represent a considerable number. As shown earlier in this report, almost 20 percent of the low-income head-of-household labor force are unemployed, amounting to nearly 100,000 potential workers.

The map in Figure 10 clearly demonstrates the clustering of low-income unemployed heads of households in the older, more densely populated sections of the region. The

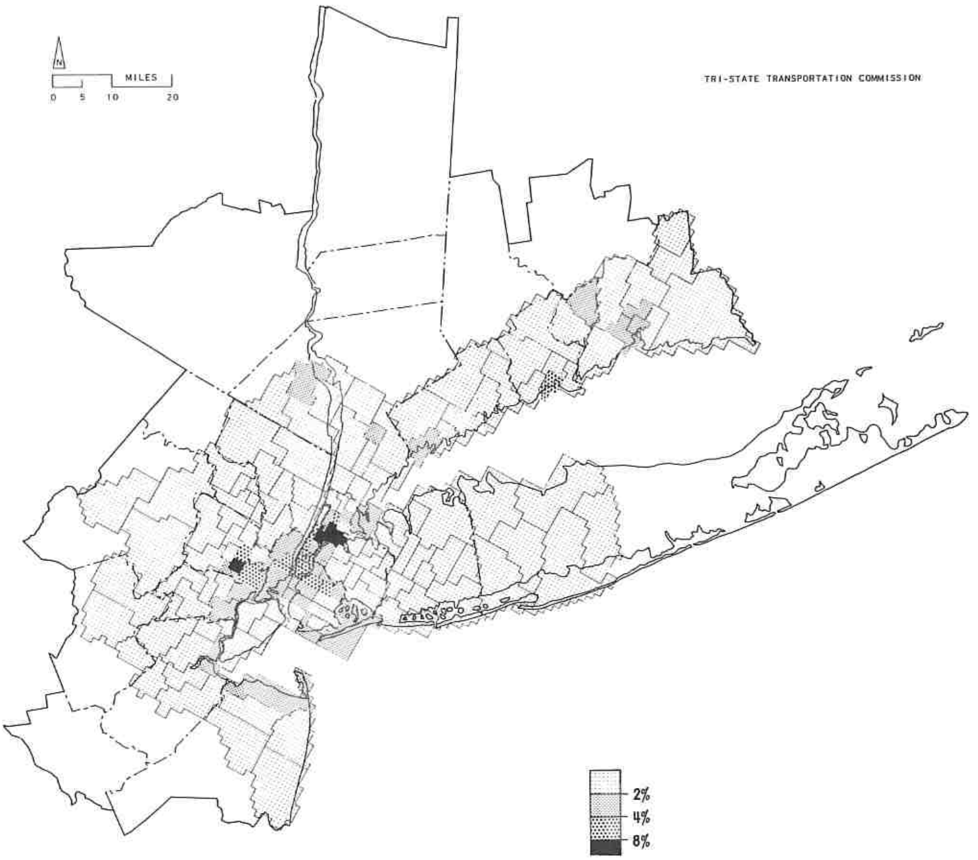


Figure 10. Percent of low-income unemployment by data aggregation district.

areas that are major centers of unemployment are in the Newark and New York City metropolitan areas, where, in the most severe cases, the rates (relative to the total labor force) range up to 8 and 13 percent respectively. Bridgeport is also a significant area of low-income unemployment.

Relative Distribution of Labor Force and Employment

The ratio of employment to resident labor force within a given area provides some indication of the amount of travel for the purpose of going to and from work required of the residents of that area. Such ratios were calculated for the Tri-State study area on a county basis (Table 14, Appendix). The results show that New York City has an overall deficit of low-income jobs for its resident labor force, whereas more employment opportunities are available in the middle- and upper-income ranges than there are residents. Within New York City, the borough of Manhattan (New York County) has the highest number of jobs in all income groups relative to its labor force, and for the most part the remaining boroughs have fewer jobs than resident workers (Table 15, Appendix). The same pattern is evident in Table 10 which gives the relative percentage of area-wide low-income labor force and employment for New York City and the rest of the study area.

TABLE 10

RELATIVE DISTRIBUTION OF LOW-INCOME LABOR FORCE AND LOW-INCOME EMPLOYMENT[a]

Location	Percent of Labor Force (1)	Percent of Low-Income Labor Force (2)	Ratio (2):(1)	Percent of Employment (3)	Percent of Low-Income Employment (4)	Ratio (4):(3)
New York City (excl. Richmond)	49.8	73.2	1.47	57.6	71.9	1.25
Outside New York City (incl. Richmond)	50.2	26.8	0.53	42.4	28.1	0.66

[a]Includes heads of households only.

LINKING HOMESITES WITH WORKSITES

The spatial link between people and jobs is transportation. The increasing mobility afforded by the private automobile to most of the population has enabled many major employers to formulate locational decisions with a declining emphasis placed on the location of the potential labor force. For most of the population, access to employment is no longer a serious constraint. For a significant minority consisting of low-income households, however, over three-quarters of whom have no private vehicles available,

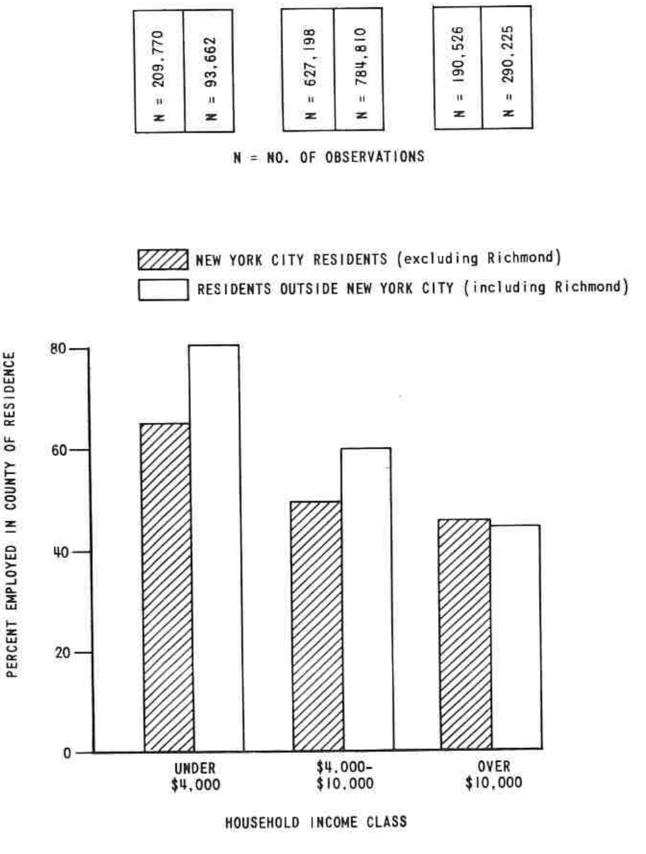


Figure 11. Percent of labor force employed in county of residence by income class.

transportation does represent a significant constraint. This section of the report examines the homesite-worksite linkage and the specific characteristics of the journey to work for low-income households in comparison with the remaining middle- and high-income population.

Percent of Labor Force Working in County of Residence

A rough measure of the work-trip mobility of the labor force is provided by examining the tendency of workers to hold jobs that are in proximity to their places of residence. In this regard, Figure 11 shows that low-income heads of households are less likely to travel outside their county of residence than those in the upper income categories. Although this relationship is preserved regardless of general residential location, it appears that low- and middle-income heads of households residing within New York City are more likely to travel outside their county of residence than those of the same groups living outside the city. This appears to be a result of both the superior transit coverage within the city and the concentration of job opportunities within Manhattan.

The combined effect of income and auto availability on the propensity to work outside the county of residence is given in Appendix Table 16. The results of this analysis show that across all income classes the availability of an automobile enhances the probability of working outside the residence county. A causal relationship is not truly demonstrated because there is no way of determining how many households forego the ownership of an automobile as a result of proximity of employment location. Given that many households do own one or more autos regardless of mode of travel to work, however, the data indicate that for low-income households with no autos available, the chance of employment outside the county of residence is relatively limited.

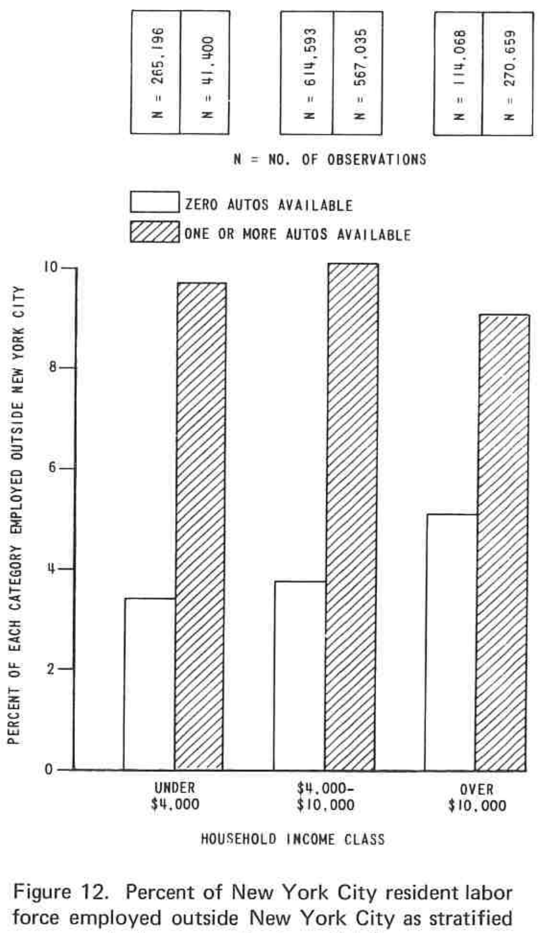


Figure 12. Percent of New York City resident labor force employed outside New York City as stratified by income and auto availability.

Reverse Commuting From New York City

The heavy concentration of low-income households residing within New York City coupled with the trend towards decentralization of employment opportunities suggested an analysis of reverse commuting from New York City. Figure 12 shows that the major factor discriminating between reverse commuters and those who remain within the city is the availability of an automobile. Regardless of income class, approximately 10 percent of all heads of households with at least one auto who reside in New York City are employed outside the city. For low-income households with zero autos this figure drops to less than 4 percent, whereas for the middle- and high-income households the absence of a private car is somewhat less of a constraint. Because the bulk of city-resident low-income households have no autos available, this group is generally

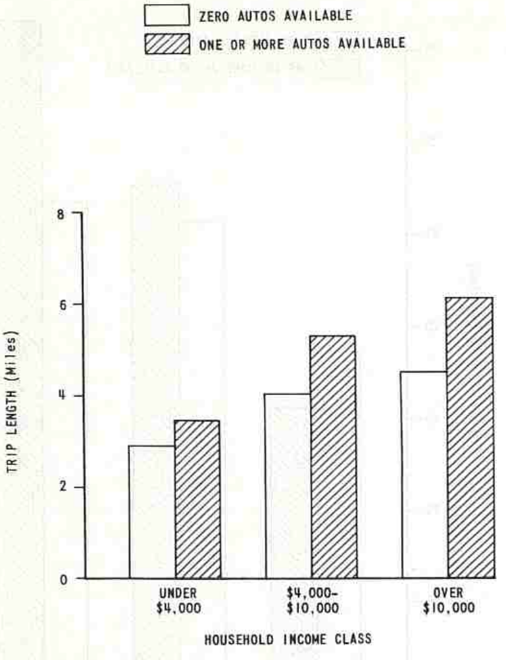


Figure 13. Average work-trip length for New York City residents.

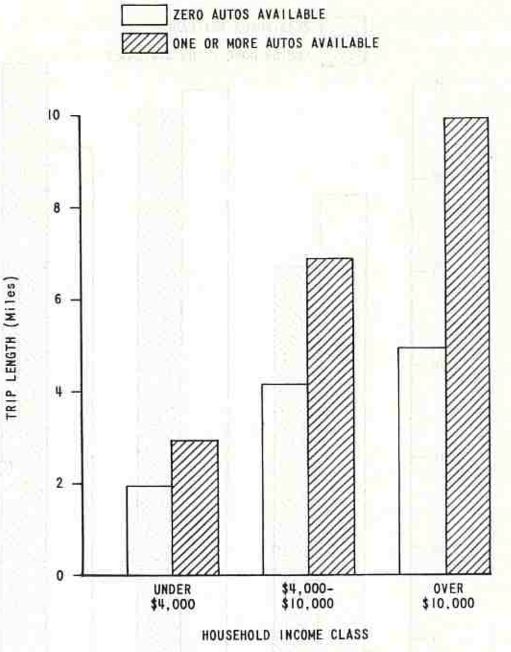


Figure 14. Average work-trip length for residents outside New York City.

much less likely to travel to work outside of New York City with the existing orientation of the transit system.

Work-Trip Length

The length of the journey to work, in terms of both distance and travel time, is a reflection of a multitude of interrelated factors. The geographical distribution of household residences with respect to centers of employment, the mode of travel used, the out-of-pocket costs incurred, and the particular occupation and industry all have a significant bearing on the time and space separation between homesites and worksites. The following discussion examines work-trip length in terms of income and auto availability and reduces the residential location bias by stratifying households according to place of residence with respect to New York City.

Distance—There is a clear correlation between distance to work and household income. Figure 13 shows that for residents of New York City the average trip length (in airline miles) for low-income heads of households is on the order of 3 miles, whereas for high-income households the average trip length is almost twice that figure. In addition, the figures for all income categories demonstrate that heads of households with one or more autos available travel longer distances between home and work than those without autos. Figure 14 shows work-trip distances for households residing outside New York City. Here, too, the shorter journey to work for low-income heads of households is substantiated; and again, those with private cars available travel greater distances.

Interestingly, the trip length for low-income households residing outside the city is shorter than for city residents whereas for high-income households the opposite is true. In the case of the former, the lack of good transit service outside New York City would tend to keep trip lengths to a minimum. For high-income households, the concentration of well-paying positions in the city, added to their ability to absorb high

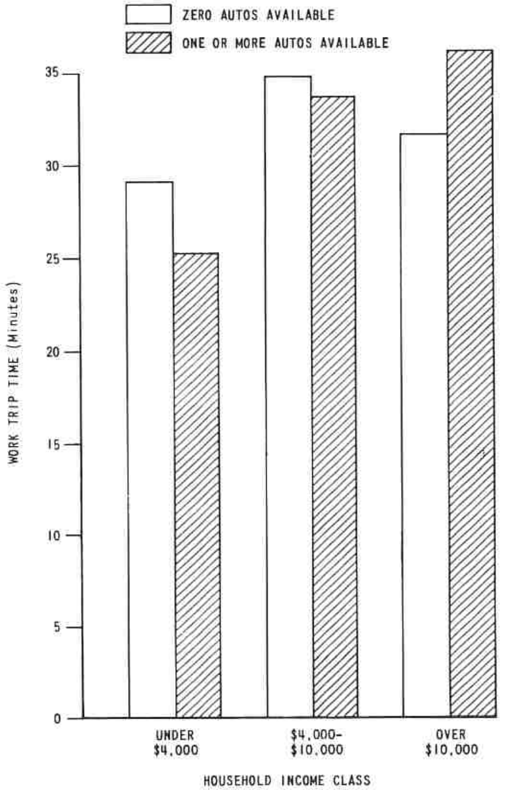


Figure 15. Average work-trip time for New York City residents.

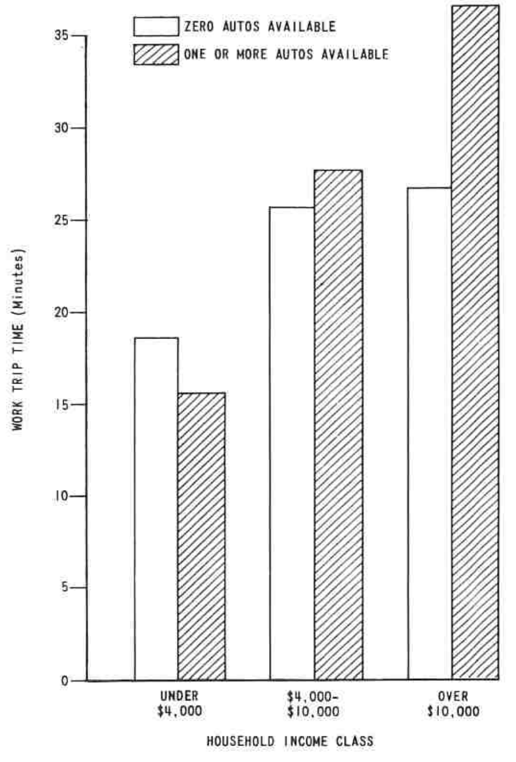


Figure 16. Average work-trip time for residents outside New York City.

travel costs, accounts for the longer journeys by nonresidents of the city.

Travel Time—Figures 15 and 16 demonstrate the relationship between income and time spent traveling to and from work for residents both inside and outside New York City. Low-income heads of households living in the city generally spend less time for their journey to work then middle- and high-income households. The addition of an auto tends to shorten the work-trip time, although as shown in the previous section, the distance traveled by those with autos is relatively longer. Regardless of auto availability, the average travel time for low-income heads of households is between 25 and 30 minutes. Middle- and high-income residents of New York City spend an average of close to 35 minutes traveling to and from work, and those in the high-income category with an auto available make longer trips in time than those with no autos.

For nonresidents of the city, the variation in work-trip time with household income is more marked than for city residents. Low-income heads of households in this subsample of the study area's population take an average of about 17 minutes to complete a journey to work whereas the mean travel time for middle-income heads is about 27 minutes and for the high-income group, well over 30 minutes. Low- and middle-income nonresidents of New York City seem to spend generally less time traveling to and from work than their city-resident counterparts. The effect of having an auto available for those living beyond the city limits varies by income class. As with city residents, low-income heads of households with a private car have a shorter work trip than those who have no private vehicle, whereas for middle- and high-income households those with autos available have a relatively longer work-trip travel time.

An indication of distance covered per 10 minutes of travel time stratified by household income, auto availability, and residential location is given in Table 11. The data indicate that low-income heads of households travel a shorter distance in a given span of time than middle- and high-income heads. Nonresidents of New York City seem to travel more swiftly than city dwellers, and the availability of an auto enhances the return on a minute's investment of travel time.

TABLE 11

AVERAGE MILES COVERED PER 10 MINUTES OF WORK-TRIP TRAVEL TIME

Residential Location	Autos Available	Mileage Stratified by Household Income		
		Under $4,000	$4,000-$10,000	Over $10,000
New York City (excl. Richmond)	0	1.0	1.2	1.4
	1+	1.4	1.6	1.7
Outside New York City (incl. Richmond)	0	1.0	1.6	1.9
	1+	1.9	2.5	3.1

There are a number of interpretations possible from the analysis of work-trip length. The results showed that low-income heads of households generally travel less in terms of distance and time than those in the middle- and high-income groups. It is not entirely clear, however, to what extent this is a result of homesite-worksite locations and to what extent it is a reflection of the more limited travel capabilities of low-income households. The stratification of households by residence and nonresidence in New York City reduced the locational bias to a degree, but obviously did not eliminate it entirely. It is likely that a combination of factors, including limited employment opportunities, reliance on mass transit, high costs of long distance commuting, and clustering of poverty-level residential centers near the older centers of both large and small urban areas, contribute to the generally shorter work-trip lengths for low-income households.

Mode of Travel

The mode of travel used for the journey to work is an important indication of the degree to which access to employment opportunities represents a problem for low-income

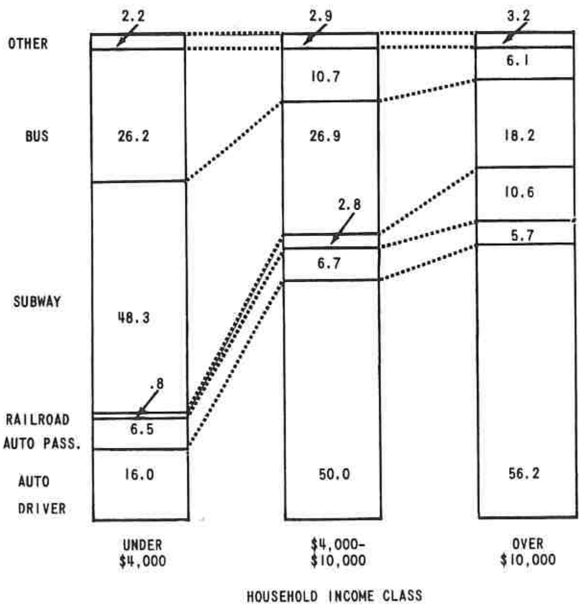


Figure 17. Percent distribution of mode for journey to work.

households. If the auto availability rate among low-income households were high, it is unlikely that transportation would be a major constraint in obtaining jobs. As shown earlier, however, relatively few low-income households have ready access to a private car. The reliance of these people upon public transportation thus is well established. The nature of the relationship between household income and the use of transit for the work trip is the subject of the following discussion.

Mode Distribution—The percent distribution of mode for the journey to work as stratified by household income is shown in Figure 17. The reliance of low-income heads of households upon mass transit modes is clearly demonstrated. About three of every four work trips made by low-income heads in the study area are via a transit mode (primarily subway and bus) whereas only slightly more than one in five are auto driver or auto passenger trips. The pattern shifts significantly in the middle- and upper-income categories where 57 and 62 percent respectively are auto driver or auto passenger trips. The growth in importance of commuter railroads as income increases is also illustrated, this being a reflection of both the outlying middle- and high-income residential areas and the ability of members of these income groups to sustain the relatively high costs associated with this travel mode.

Use of Mass Transit—In analyzing mass transit usage by the low-, middle-, and high-income groups, homesites and worksites were classified with respect to location within or outside New York City. For New York City residents there is a clear tendency toward declining transit patronage as household income increases, regardless of

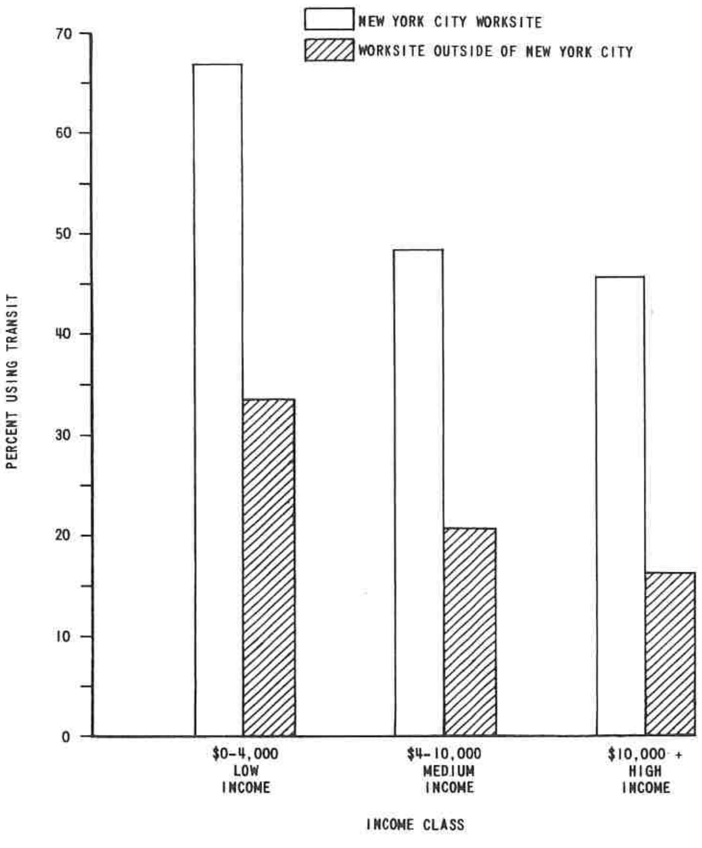


Figure 18. Use of mass transit for journey to work by New York City residents.

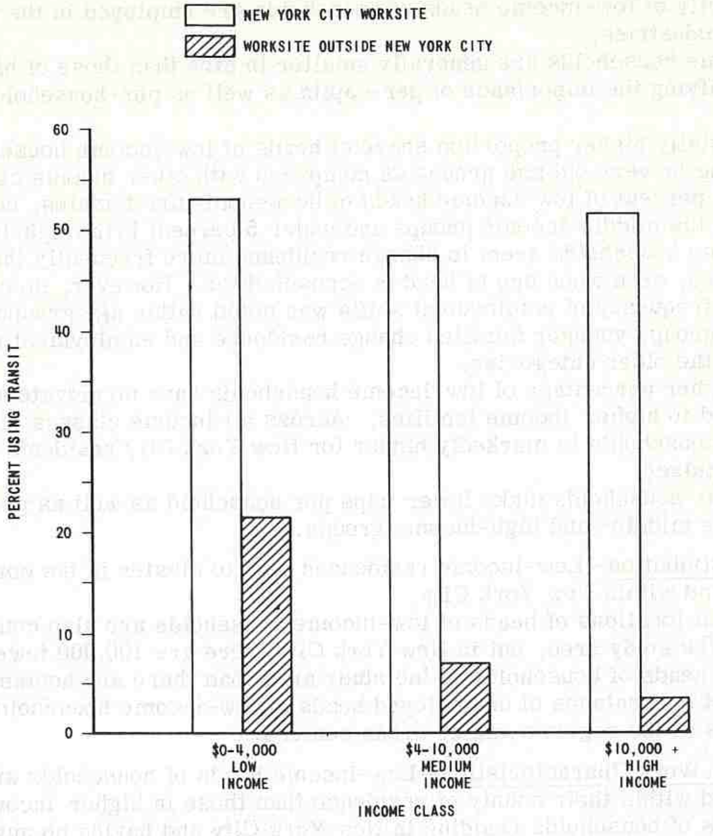


Figure 19. Use of mass transit for journey to work by residents outside New York City by income and worksite.

work location (Fig. 18). In addition, transit usage by residents employed within the city is about twice that for residents working outside the city regardless of income, reflecting in part the service, coverage, and orientation of the New York City transit system.

For residents outside New York City, the use of transit by those employed in New York City is fairly constant at around 50 percent. For those both residing and working outside the city, however, transit usage varies from 20 percent for the low-income group to under 5 percent for those in the highest income category (Fig. 19).

SUMMARY OF RESULTS

It was the purpose of this study to provide a profile of some of the major characteristics of low-income households that have some bearing on the relationship between transportation and employment opportunities in the Tri-State study area. The analyses performed were centered about factors that are related to the annual income of the family unit and that reflect to some extent the relative degree to which transportation remains a critical constraint to poverty-level heads of households in locating jobs. Some of the important findings contained within this report are listed below.

Major Findings

Socioeconomic Characteristics—Almost 20 percent of low-income heads of households in the labor force are out of work. This amounts of 100,000 men and women.

The majority of low-income heads of households are employed in the manufacturing and service industries.

Low-income households are generally smaller in size than those of higher income groups, signifying the importance of per-capita as well as per-household income measures.

A substantially higher proportion share of heads of low-income households are in the very young or very old age groups as compared with other income classes.

Almost 40 percent of low-income heads of households are females, compared with 10 percent in the middle-income groups and under 5 percent in the high-income groups.

Low-income households seem to change residence more frequently than higher-income familes, even when age of head is accounted for. However, no clear relationship between frequency of employment shifts was noted within age groups. Within each income group, younger families change residence and employment more often than those in the older categories.

A much higher percentage of low-income households have no private vehicles available compared to higher income families. Across all income classes the percentage of zero-auto households is markedly higher for New York City residents than for the remaining populace.

Low-income households make fewer trips per household as well as per person than families in the middle- and high-income groups.

Spatial Distribution—Low-income residences tend to cluster in the core of the study area around and within New York City.

Employment locations of heads of low-income households are also concentrated in the center of the study area, but in New York City there are 100,000 fewer low-income worksites for heads of households in the study area than there are homesites.

The highest percentages of unemployed heads of low-income households occur in the core areas of the region's major urban centers.

Journey-to-Work Characteristics—Low-income heads of households are more likely to be employed within their county of residence than those in higher income families.

Of all heads of households residing in New York City and having no auto available, low-income households are the least likely to be employed outside of the city. Compared with these heads of households, those with one or more autos are at least twice as likely to reverse commute from New York City regardless of income.

Low-income heads of households work closer to home than the rest of the population. Regardless of income, those with an auto available make longer trips.

Low-income heads of households spend, on the average, less time traveling to or from work than those in the higher income groups.

The use of mass transit for the work trip is generally much more predominant among low-income heads of households than among those in the middle- and high-income groups.

Future Research

The research reported in this study is of a descriptive nature. It characterizes the composition of low-income households by means of demographic data and identifies their residential and worksite location and the transportation link between home and worksite. The effect of transportation on poverty may only be implied in this study through correlative measures associating auto ownership, transit usage, population, and employment density with job opportunities and employment status.

Future research should take a more concentrated analytical approach to the problems of poverty. This will come about only with an increased awareness and knowledge of the dimensions of the problem by all disciplines. In the field of transportation planning, sophisticated analytical tools are available to synthesize travel behavior, including the journey to work. Residential and employment growth have also been simulated in comprehensive computer modeling efforts. Efforts should be initiated to use these analytical models in testing the various effects of alternate approaches to the transportation problems of the poor.

In addition to analytically simulating the residence-employment relationship of low-income households, it is essential to have a much more enriched data source available to describe low-income households. The Tri-State Home Interview Survey permitted a multidimensional view of the household structure through specified cross-classifications. It was limited, however, because it was only a 1 percent sample of the entire population and was not designed specifically for this type of study.

One approach used to test methods of relieving the transportation problems of the poor is via mass transportation demonstration projects financed in large part by the federal government. These projects have the advantage of measuring the cause-effect relationships between transportation and poverty on a real-life dimension. Examples of such demonstration projects include the Watts Project and the Nassau-Suffolk study. In the Watts Project, the effect of improved transit service was analyzed by its impact on employment and other opportunities for the area's residents; and the Nassau-Suffolk study is testing whether an improved accessibility linkage between labor supply and demand in areas outside the central business district would increase employment opportunities among unskilled and semiskilled workers. Similar projects and studies are under way in several other major urban areas.

Perhaps the most effective means of understanding the link between transportation and poverty is the combined use of more meaningful data, analytical tools that can aid in testing alternate solutions, and demonstration projects that can apply these solutions under real-life circumstances.

REFERENCE

1. Fisher, R. J., and Sosslau, A. B. Census Data as a Source for Urban Transportation Planning. Highway Research Record 141, 1966, pp. 53-56.

Appendix

DATA STRATIFIED BY COUNTY OF RESIDENCE

Data pertaining to labor force, employment, income, and automobile availability are given in Tables 12, 13, 14, 15, and 16.

TABLE 12

PERCENT DISTRIBUTION OF LABOR FORCE[a] WITHIN EACH COUNTY[b] BY INCOME AND AUTO AVAILABILITY

County	Unemployed (percent) (Under $4,000)		Employed (percent)						Total Labor Force[c]
			Under $4,000		$4,000-$10,000		Over $10,000		
	0 Auto	1+ Auto	0 Auto	1+ Auto	0 Auto	1+ Auto	0 Auto	1+ Auto	
New York	5.3[d]	0.3	21.0	1.7	39.4	11.1	10.7	8.6	577,400
Bronx	3.4	0.3	11.1	1.9	32.0	32.4	4.1	13.2	358,500
Kings	2.7	0.3	13.1	3.1	29.9	35.5	2.8	10.6	680,900
Queens	0.8	0.1	4.4	1.6	18.6	42.4	3.8	27.1	494,700
N.Y.C. (excl. Richmond)	3.0	0.3	13.0	2.2	30.3	29.8	5.5	14.3	2,111,500
Richmond	0.2	0.2	3.7	3.3	12.9	60.5	0.2	17.8	54,700
Nassau	0.3	0.2	0.9	1.3	2.9	50.2	0.6	42.5	323,700
Suffolk	0.1	0.1	0.7	1.4	2.5	62.6	0.4	30.4	147,000
Westchester	0.5	0.4	3.5	1.7	8.4	40.3	1.0	42.4	188,300
Rockland	0.6	0.3	1.8	1.5	1.2	64.5	0.0	28.8	36,800
Hudson	2.4	0.8	5.5	3.3	20.0	48.7	1.7	15.4	154,800
Essex	2.4	0.6	6.1	3.7	11.5	49.7	0.6	23.5	239,000
Bergen	0.2	0.3	1.2	1.9	3.2	52.6	1.0	38.3	210,100
Passaic	0.7	0.2	5.8	3.7	6.4	59.2	0.3	21.7	96,600
Morris	0.0	0.2	1.3	1.3	2.1	61.5	0.0	33.1	55,300
Union	0.9	0.6	1.8	1.4	5.0	58.8	0.2	30.6	134,200
Somerset	0.0	0.3	1.0	2.0	2.3	64.3	0.0	30.0	32,600
Middlesex	0.9	0.3	1.3	2.9	3.0	71.2	0.1	18.7	106,300
Monmouth	0.6	0.7	2.9	5.4	2.1	63.0	0.0	23.2	78,000
S.W. Conn.	0.4	0.9	3.2	2.3	3.5	44.0	0.7	44.0	78,900
Bridgeport	1.1	0.5	3.3	3.5	6.7	62.3	0.3	21.4	69,200
Ansonia Derby	1.8	0.0	2.5	4.3	4.9	62.6	0.0	22.7	16,500
South Central	0.6	0.6	2.6	3.2	5.0	67.2	0.4	19.6	110,100
Outside N.Y.C. (incl. Richmond)	0.8	0.4	2.9	2.5	6.5	54.8	0.6	30.1	2,132,100
Within Cordon	1.9	0.3	7.9	2.4	18.4	42.3	3.0	22.2	4,243,600

[a]Percentages across do not total 100.0 because unemployed reporting over $4,000/year are excluded from this tabulation.
[b]For counties divided by the Cordon, figures apply to portion lying within the Cordon.
[c]Includes heads of households only.
[d]Table is read as follows: 5.3 percent of all heads of households in the labor force residing in New York County (Manhattan) are unemployed, are members of households reporting income under $4,000/year, and have no private autos available.

TABLE 13

PERCENT DISTRIBUTION OF EMPLOYMENT[a] WITHIN EACH COUNTY[b] BY INCOME CLASSIFICATION

County	Employed (percent)			Total Employed
	Under $4,000	$4,000-$10,000	Over $10,000	
New York	13.0[c]	57.0	30.0	1,370,600
Bronx	13.9	64.6	21.5	172,400
Kings	16.7	64.8	18.5	456,000
Queens	10.5	64.8	24.7	329,000
N.Y.C. (excl. Richmond)	13.4	60.2	26.4	2,328,000
Richmond	9.6	70.3	20.1	31,700
Nassau	5.3	62.2	32.5	213,400
Suffolk	3.1	67.1	29.8	95,200
Westchester	8.8	59.6	31.6	134,700
Rockland	6.3	68.8	24.9	23,500
Hudson	9.0	69.7	21.3	167,200
Essex	8.5	66.9	24.6	237,600
Bergen	6.1	65.2	28.7	146,800
Passaic	9.9	65.7	24.4	90,300
Morris	4.9	65.8	29.3	45,500
Union	5.5	70.3	24.2	123,300
Somerset	5.5	72.9	21.6	25,300
Middlesex	4.5	74.6	20.9	85,000
Monmouth	12.5	67.5	20.0	53,600
S.W. Conn.	7.8	63.2	29.0	63,200
Bridgeport	6.4	70.3	23.3	70,300
Ansonia Derby	9.1	65.7	25.2	11,000
South Central	6.7	74.9	18.4	97,300
Outside N.Y.C. (incl. Richmond)	7.1	67.1	25.8	1,714,900
Total	10.7	63.1	26.2	4,042,900

[a]Includes heads of households only.
[b]For counties divided by the Cordon, figures apply to portion lying within the Cordon.
[c]Table is read as follows: 13.0 percent of all heads of households employed in New York County (Manhattan) report household incomes under $4,000/year.

TABLE 14

COUNTY[a] RATIOS OF EMPLOYMENT[b] TO LABOR FORCE[b] BY HOUSEHOLD INCOME CLASSIFICATION

County	Employed:Resident Labor Force			All Income Classes
	Under $4,000	$4,000-$10,000	Over $10,000	
New York	1.08[c]	2.62	3.61	2.37
Bronx	0.40	0.47	0.59	0.48
Kings	0.58	0.64	0.93	0.67
Queens	1.02	0.69	0.53	0.67
N.Y.C. (excl. Richmond)	0.81	1.07	1.46	1.10
Richmond	0.74	0.55	0.64	0.58
Nassau	1.28	0.76	0.50	0.66
Suffolk	0.86	0.65	0.63	0.65
Westchester	0.97	0.84	0.53	0.72
Rockland	0.95	0.66	0.54	0.64
Hudson	0.80	1.07	1.33	1.08
Essex	0.63	1.05	1.05	0.99
Bergen	1.19	0.53	0.51	0.70
Passaic	0.87	0.91	1.03	0.94
Morris	1.38	0.84	0.73	0.82
Union	1.04	1.01	0.72	0.92
Somerset	1.18	0.85	0.56	0.78
Middlesex	0.65	0.79	0.88	0.80
Monmouth	0.85	0.70	0.59	0.69
S.W. Conn.	0.91	1.05	0.52	0.80
Bridgeport	0.76	1.02	1.09	1.02
Ansonia Derby	0.71	0.64	0.74	0.67
South Central	0.84	0.91	0.81	0.88
Outside N.Y.C. (incl. Richmond)	0.85	0.83	0.67	0.80
Total	0.82	0.95	0.98	0.95

[a]For counties divided by the Cordon, figures apply to portion lying within the Cordon.

[b]Includes heads of households only.

[c]Table is read as follows: The number of low-income heads of households employed in New York County exceeds the low-income resident labor force in New York County by 8 percent.

TABLE 15

RELATIVE COUNTY PERCENTAGES OF LOW-INCOME LABOR FORCE AND EMPLOYMENT[a]

County	Percent of Labor Force (1)	Percent of Low-Income Labor Force		Ratios		Percent of Employment (4)	Percent of Low-Income Employment (5)	Ratio (5):(4)
		Employed (2)	Unemployed (3)	(2):(1)	(3):(1)			
Manhattan	13.6	30.6	32.9	2.25	2.42	33.9	41.1	1.21
Bronx	8.4	10.6	13.7	1.26	1.63	4.3	5.5	1.28
Brooklyn	16.0	25.5	21.0	1.59	1.31	11.3	17.6	1.56
Queens	11.7	6.8	4.5	0.58	0.38	8.1	8.0	0.99
N.Y.C. (excl. Richmond)	49.7	73.5	72.1	1.48	1.45	57.6	72.2	1.25
Richmond	1.3	0.9	0.2	0.69	0.15	0.8	0.7	0.88
Nassau	7.6	1.6	1.8	0.21	0.24	5.3	2.6	0.49
Suffolk	3.5	0.7	0.3	0.20	0.09	2.4	0.2	0.08
Westchester	4.4	2.4	1.9	0.55	0.43	3.3	2.7	0.82
Rockland	0.9	0.3	0.3	0.33	0.33	0.6	0.3	0.50
Hudson	3.6	3.2	5.2	0.89	1.44	4.1	3.5	0.85
Essex	5.6	5.7	7.8	1.02	1.39	5.9	4.8	0.81
Bergen	5.0	1.5	0.9	0.30	0.18	3.6	2.1	0.58
Passaic	2.3	2.2	1.0	0.96	0.43	2.2	2.1	0.91
Morris	1.3	0.3	0.1	0.23	0.08	1.1	0.5	0.45
Union	3.2	1.0	2.1	0.31	0.66	3.1	1.7	0.55
Somerset	0.8	0.2	0.1	0.25	0.12	0.6	0.3	0.50
Middlesex	2.5	1.0	1.4	0.40	0.56	2.1	0.9	0.43
Monmouth	1.8	1.6	1.0	0.89	0.56	1.3	1.6	1.23
S.W. Conn.	1.9	1.0	1.1	0.3	0.58	1.6	1.1	0.69
Bridgeport	1.6	1.1	1.1	0.69	0.69	1.7	1.0	0.59
Ansonia Derby	0.4	0.3	0.3	0.75	0.75	0.3	0.2	0.67
South Central	2.6	1.5	1.3	0.58	0.50	2.4	1.5	0.63
Outside N.Y.C. (incl. Richmond)	50.3	26.5	27.9	0.53	0.55	42.4	27.8	0.66

[a]Includes heads of households only.

TABLE 16

PERCENT OF LABOR FORCE[a] IN EACH OF SIX INCOME-AUTO AVAILABILITY CATEGORIES EMPLOYED IN COUNTY[b] OF RESIDENCE

County	Resident Labor Force (percent)						All Classes
	Under $4,000		$4,000-$10,000		Over $10,000		
	0 Auto	1+ Auto	0 Auto	1+ Auto	0 Auto	1+ Auto	
New York	84.7[c]	70.7	83.2	73.7	86.1	80.9	82.3
Bronx	30.3	44.2	25.8	35.7	18.1	31.0	30.5
Kings	58.3	67.7	45.5	49.9	32.5	45.6	49.3
Queens	59.6	51.6	32.7	36.9	22.2	27.3	34.1
N.Y.C. (excl. Richmond)	66.2	62.1	53.5	45.3	58.0	41.0	51.2
Richmond	84.4	64.6	38.0	49.5	100.0	49.6	49.9
Nassau	89.1	74.4	59.2	47.2	65.5	33.5	42.5
Suffolk	70.1	68.7	74.1	50.2	48.6	44.4	49.4
Westchester	83.9	93.2	72.8	61.5	60.6	39.9	54.7
Rockland	83.6	100.0	74.8	54.1	—	34.7	50.0
Hudson	80.3	84.7	63.1	63.9	65.0	57.3	64.4
Essex	74.4	71.5	76.1	65.9	61.2	54.9	65.4
Bergen	91.1	63.1	36.8	48.5	30.3	36.3	44.0
Passaic	81.9	72.3	80.3	51.3	32.2	47.7	55.1
Morris	71.4	100.0	82.3	57.1	—	46.0	54.8
Union	81.7	78.7	77.6	53.4	100.0	42.6	52.3
Somerset	100.0	85.2	56.6	54.0	—	31.7	48.6
Middlesex	67.4	72.6	79.7	56.4	100.0	51.3	56.9
Monmouth	100.0	89.9	87.3	65.1	—	55.6	66.0
S.W. Conn.	96.1	88.5	96.2	86.1	80.0	43.9	67.8
Bridgeport	90.7	91.3	86.4	83.7	100.0	82.4	84.2
Ansonia Derby	100.0	71.3	86.5	51.9	—	45.7	54.2
South Central	96.2	94.2	98.1	85.1	100.0	76.1	84.7
Outside N.Y.C. (incl. Richmond)	82.7	79.2	69.7	59.0	59.9	44.1	56.3
Total	69.3	71.4	56.3	54.2	58.2	43.1	53.8

[a]Includes heads of households only.

[b]For counties divided by the Cordon, figures apply to portion lying within the Cordon.

[c]Table is read as follows: 84.7 percent of all employed heads of households who reside in New York County (Manhattan), whose reported household income is under $4,000/year, and who have no private vehicles available, are employed in New York County.

Urban Travel and City Structure

ALAN M. VOORHEES and SALVATORE J. BELLOMO, Alan M. Voorhees and Associates, Inc.

Urban travel as measured by the length of the work trip (in miles and minutes) was found to be highly related to the structure of a city, as reflected by the time distribution of job opportunities within a metropolitan area. This measure of city structure combines effects of the spatial allocation of jobs and the speed of the transportation network. It is clear from research of this relationship that, as the transportation network has been improved through an increase in network speed, greater mobility has been afforded to the increasing population, which has allowed for new kinds of development at lower residential densities, and the work-trip length in miles has increased. As mobility has increased, the average trip length of job opportunities has increased. It would thus appear that the selection of city structure and of broader environmental and living preference objectives are the key decisions that should be made in a metropolitan area. Once city structure and environmental objectives are selected, care must be exercised by the planner to develop a transportation system directed toward these objectives.

•THE INTERRELATIONSHIP between urban travel and city structure has often been discussed. The early works of Mitchell (1) certainly demonstrated that there was a strong tie between city structure and urban travel. It is only recently, however, that this complex interrelationship is beginning to be understood; and it seems appropriate now to examine what this research is indicating so that cities and associated transportation systems can be better planned.

URBAN TRAVEL

In the past, various techniques to quantify urban travel have been developed. This has largely been done by breaking travel into trip purposes, such as work, shop, and recreation. It is also advantageous to look at travel for these purposes in terms of average trip length (in minutes or miles) or trip-length distribution (2).

Research has indicated that the most frequent trip is the work trip and that this trip is also the longest made on the average, in many cases accounting for 40 percent of the vehicle-miles traveled in a metropolitan area. The other trips are important as well, particularly the social-recreation trips, which seem to be growing at a faster rate than all others. Planners will have to consider these trips more when planning transportation and land-use systems.

CITY STRUCTURE

In dealing with the urban structure, there has been no effective way of describing the socioeconomic structure of the city. Considerable work, however, has been done in describing the city in terms of density versus distance from the central business district (CBD), as in the early work of Colin Clark (3). As shown by Figure 1, the population density descends rapidly as distance from the CBD increases. The population density near the core is much higher in the older cities, such as Philadelphia and St. Louis, than it is in a city like Houston, which has developed largely in the auto age.

Paper sponsored by Committee on Origin and Destination.

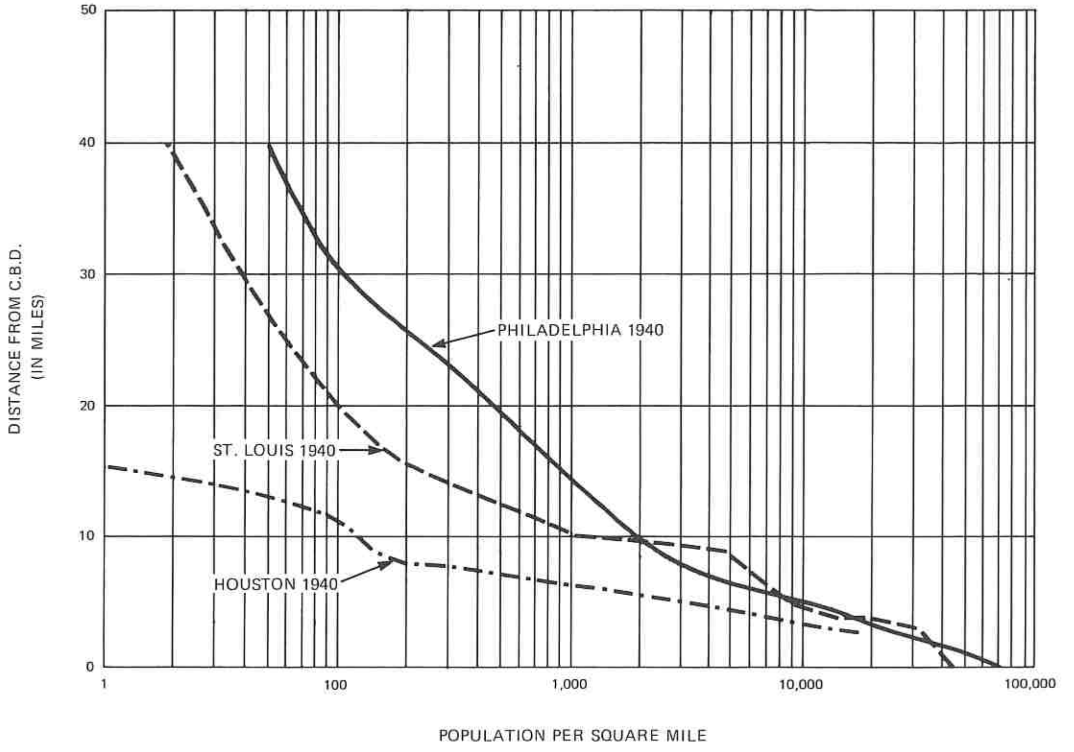


Figure 1. Residential density patterns for different metropolitan areas. (From Arthur Row and Ernest Jurrat. The Economic Forces Shaping Land Use Patterns. Jour. Amer. Inst. Planners, Vol. 25, No. 2, p. 78.)

There is variation within the general framework shown in Figure 1. A more recent way to measure the socioeconomic structure of a city in a more precise manner is to develop a distribution for work-trip opportunities. This appears to be a better way of stating the structure of a city. In effect, by this technique we measure the distance from a residential area to all the jobs in the area. This can be presented as a distribution of trip opportunities (Fig. 2). It can also be stated for various sections of the metropolitan area, or it can be aggregated into a mean value for the metropolitan area.

An advantage of this approach is that changes in the structure of a city can be determined. Figure 2 is a good illustration of this; it shows the distribution change in Washington between 1948 and 1955 which occurred largely with no substantial change in the speed of the transportation network.

VARIATIONS IN TRIP LENGTH

Residential density patterns for the Washington metropolitan area (Fig. 3) show the typical declining residential density profile from the CBD. The work-trip opportunity distribution related to an area close to the CBD has characteristics different from those in an area farther away. These differences can be demonstrated by taking three zones within the Washington area at different distances from the downtown area. These zones are labeled A, B, and C and are 8, 14, and 21 minutes from the downtown area respecitvely.

The opportunity distribution related to these three zones is shown in Figure 4. As one would expect, zone A has greater job opportunities with 5- to 10-minute trips because it is closer to the job opportunities, particularly those in the downtown area. Zone B has most of its job opportunities with 10- to 15-minute trips, whereas zone C

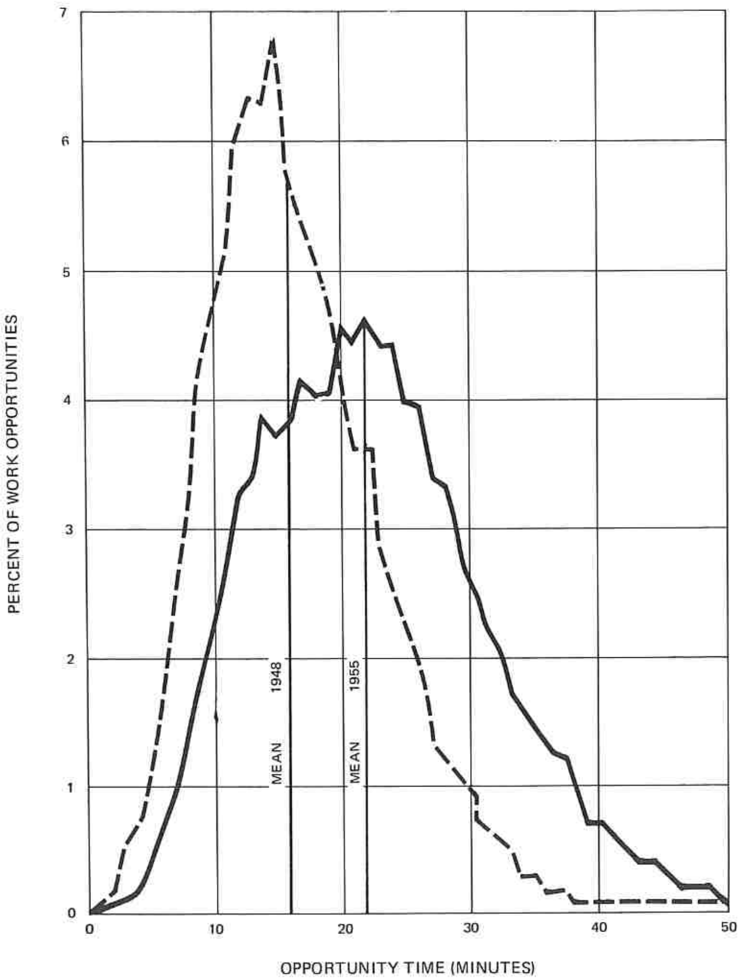


Figure 2. Opportunity distribution for Washington, D. C., 1948/1955. (From Alan M. Voorhees, Salvatore J. Bellomo, Joseph L. Schofer, and Donald E. Cleveland. Factors in Work Trip Length. Highway Research Record 141, 1966, pp. 24-46.)

has the bulk of its job opportunities within the 20- to 25-minute trip range. The work-trip lengths that are found for these zones reflect these differences in the opportunity distribution. As shown in Figure 5, the trip distribution patterns are related to the opportunity distribution. The average trip in zone A is smaller than those in zones B and C, and zone A trips are concentrated in the shorter time ranges. Trips for B and C are longer in length. This relationship can be estimated through the use of the gravity model trip distribution formula. There is a direct relationship between the opportunity trip length and the average trip length. Figure 6 shows this relationship for the three zones analyzed.

CITY STRUCTURE AND TRIP LENGTH

The total trip pattern that is produced in a metropolitan area is the composite of all the opportunity distributions in various sections of the metropolitan area; it is not

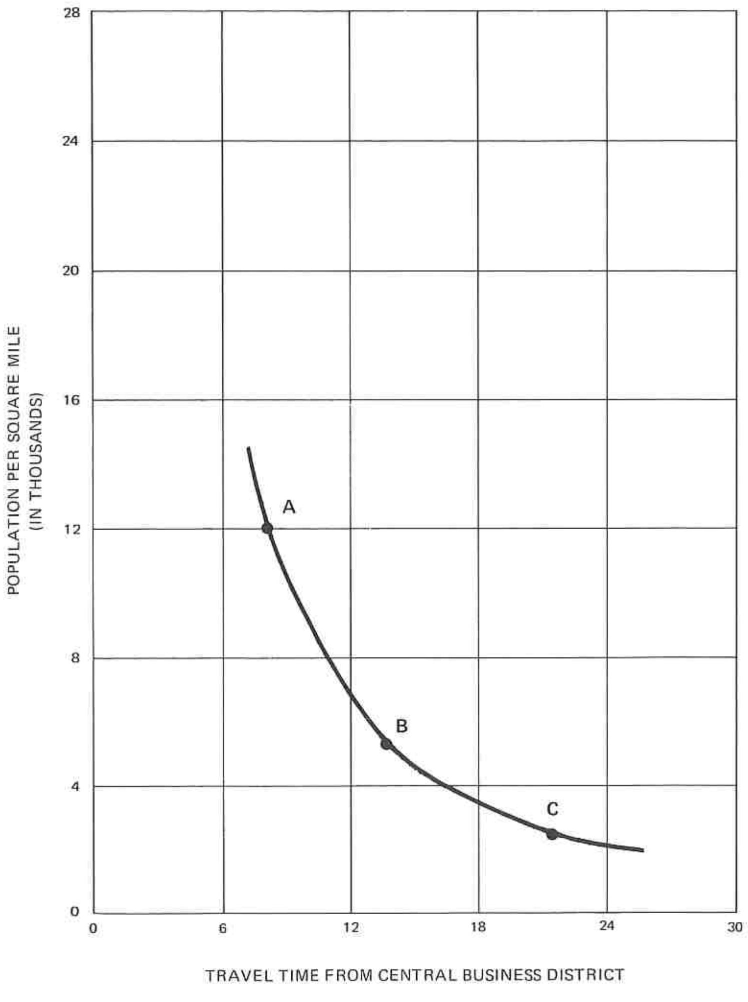


Figure 3. Residential density profile for Washington, D. C. (From Economic Base Study for the General Development Plan—National Capital Region. Council for Economic and Industry Research, Inc., Washington, D. C., June 1956.)

surprising, therefore, that the same kind of relationship can be found for the overall opportunity distribution across various cities. Figure 7 shows the work-trip opportunity distributions for three cities that have quite marked differences in their structure. Aside from the size differences, which affect the work-trip opportunity distribution, the cities represent quite different structures. In Erie, most of the trip opportunities are within 20 minutes; in Detroit, these work-trip opportunities are largely less than 40 minutes long; and in Seattle-Tacoma the trip opportunity distribution seems to be almost flat. It is not surprising, therefore, that we find the cities of Seattle and Tacoma with the longest actual work-trip lengths and Erie with the shortest because of this difference in city structure. This relationship between trip length and the opportunity length is shown in Figure 8. It was found that the work-trip length increased as the opportunity length increased in several cities and as a result of the simulation study.

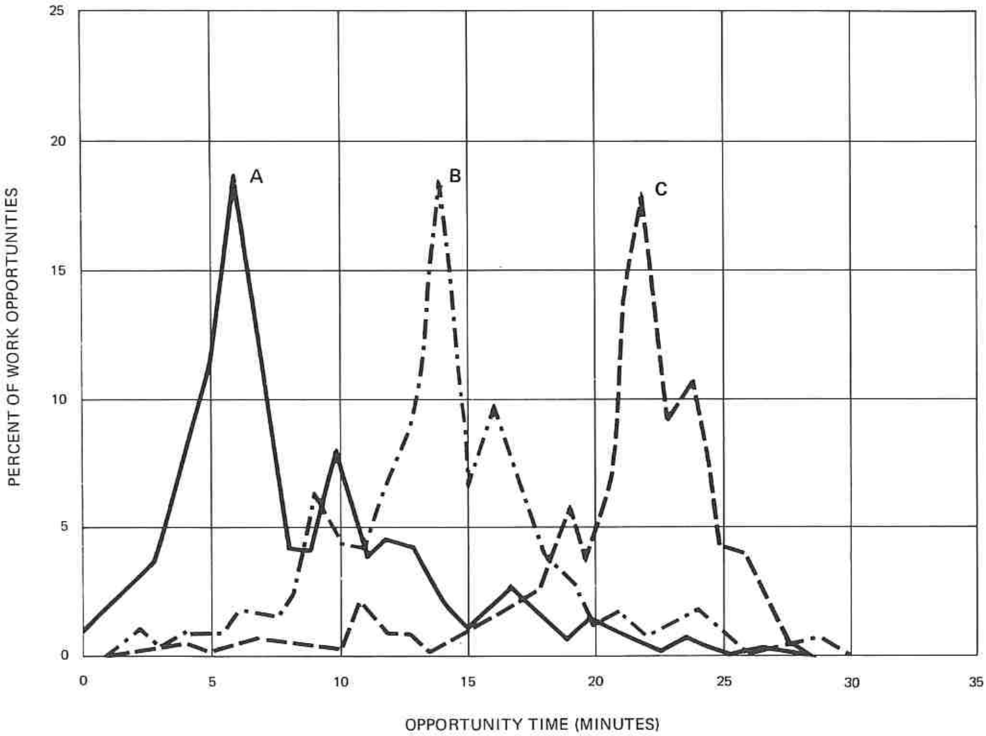


Figure 4. Opportunity distributions for selected zones in Washington, D. C., 1948. (See Fig. 2 for source.)

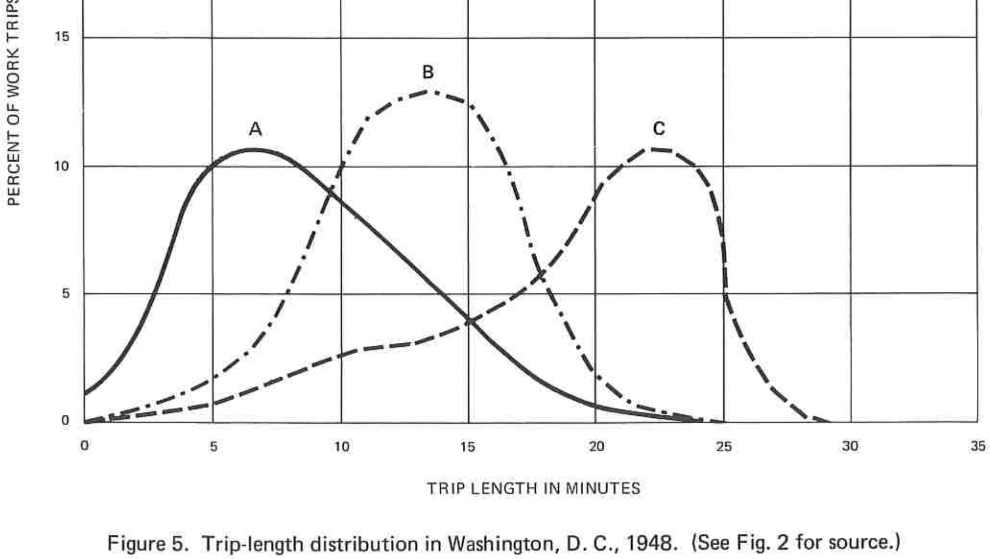


Figure 5. Trip-length distribution in Washington, D. C., 1948. (See Fig. 2 for source.)

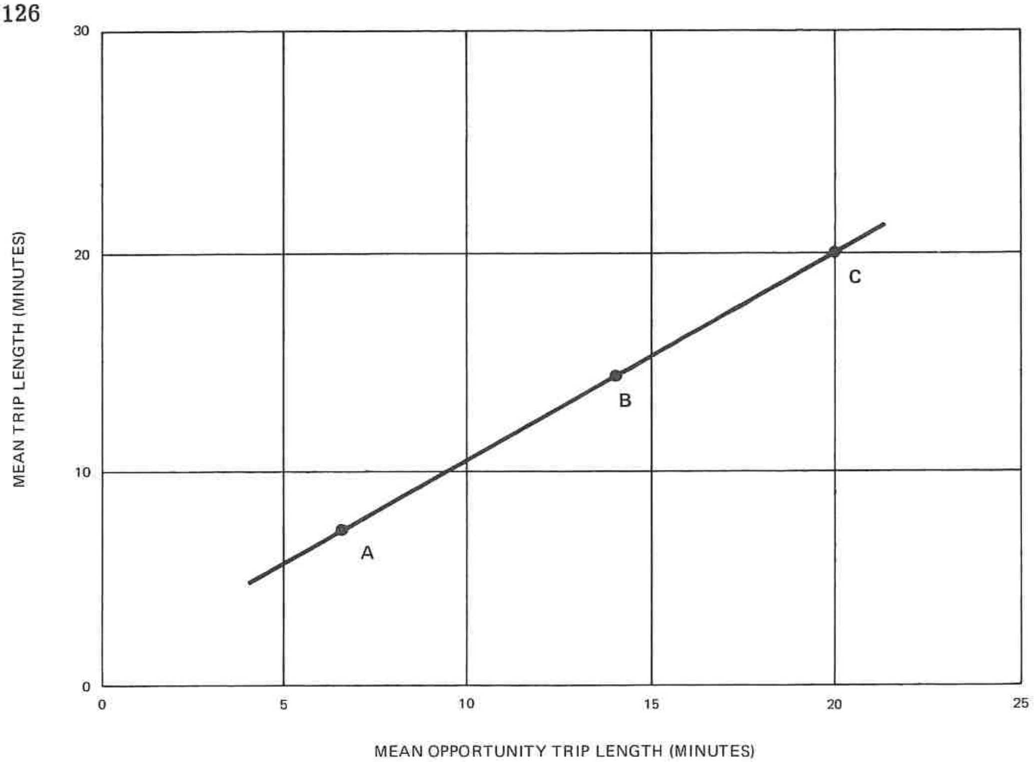


Figure 6. Mean work-trip length versus mean opportunity-trip length. (See Fig. 2 for source.)

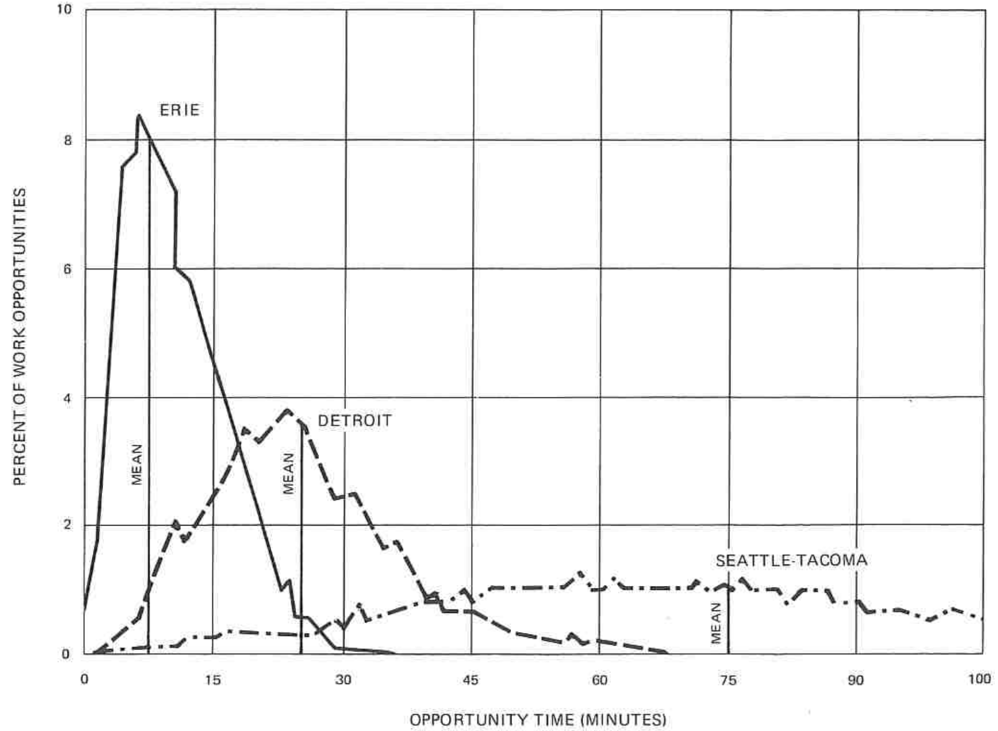


Figure 7. Opportunity distribution across three cities. (See Fig. 2 for source.)

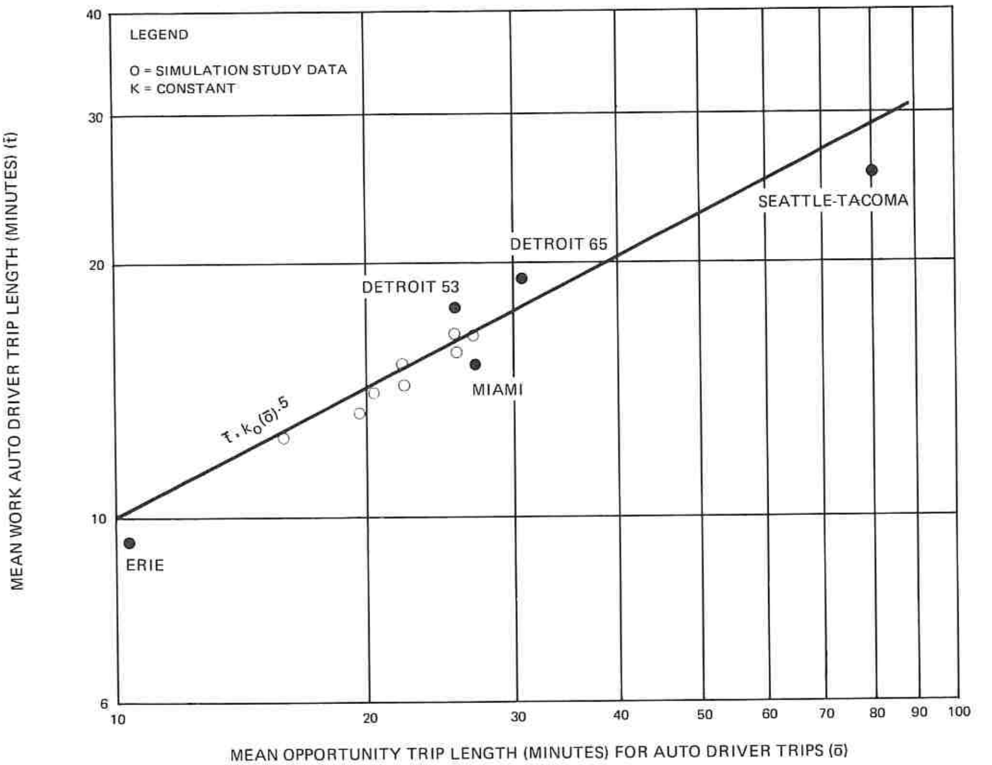


Figure 8. Mean work-trip length versus mean opportunity-trip length. (From Alan M. Voorhees and Salvatore J. Bellomo. Factors and Trends in Trip Lengths. National Cooperative Highway Research Project 7-4, 1969.)

NETWORK SPEED AND TRIP LENGTH

Now that we have established the relationship between city structure, as measured by the opportunity trip-length distribution with its associated mean value, and urban travel, as reflected by the work-trip distribution, we can analyze what happens when the speed of a transportation network is increased. If we take a simple case, for example, where the travel time to all the other zones is reduced by increasing the speed by 50 percent, the opportunity distribution shown in Figure 9 actually changes. If we apply an average set of travel time factors for the work trip, the trip length will change. In fact, the trip length in terms of minutes actually reduces slightly, whereas the trip length in miles increases. Various simulation studies of the work trip have demonstrated that trip length in miles for a constant urban form is proportional to about the 0.75 power of the change in network speed, whereas change in trip length in minutes is inversely proportional to about the 0.25 power of change in network speed.

The relationship between the opportunity trip length and average network speed can also be derived mathematically. We know from Figure 8 that a relationship exists between the mean work-trip length (in minutes) and the mean opportunity work-trip length. This relationship, which was verified using cross-sectional, historical, and simulation study data, can be stated as

$$t = k_1 O^{0.5}$$

where

t = mean work-trip length, minutes;
O = mean opportunity trip length, minutes; and
k_1 = a constant.

For two points in time (t_2 and t_1) this relationship would expand to

$$\frac{t_2}{t_1} = \frac{O_2^{0.5}}{O_1} \tag{1}$$

We also know from the historical and cross-sectional research analyses in Detroit, Reading, Washington, and so forth that

$$L = k_2 P^{0.2} S^{1.5}$$

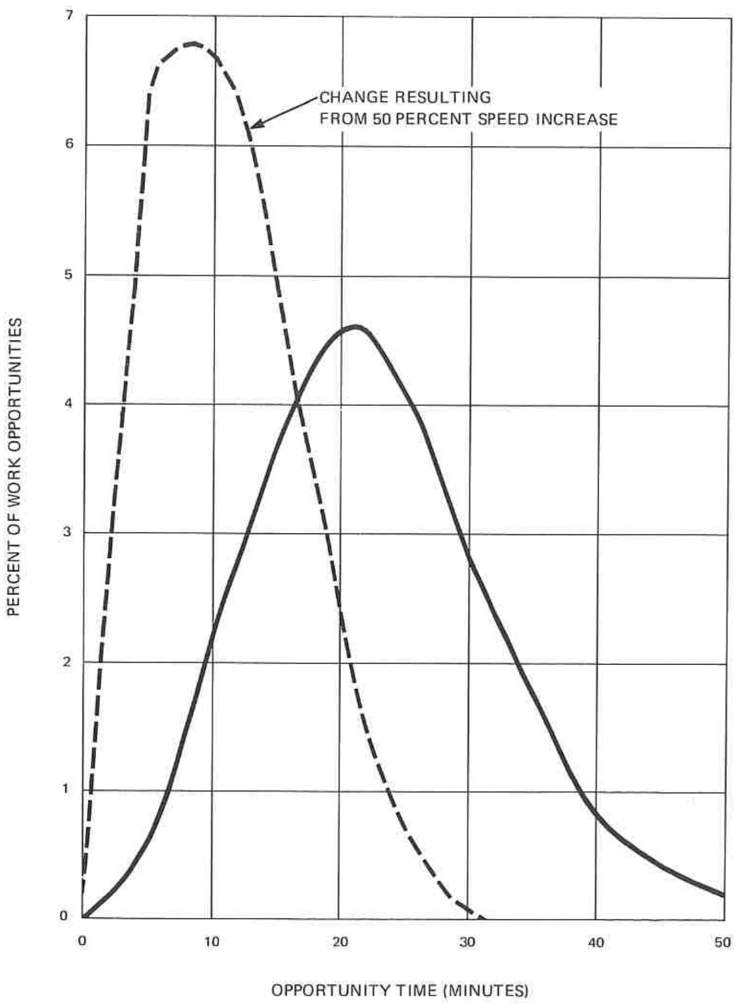


Figure 9. Opportunity distribution change developed by change in network speed.

where

L = mean work-trip length, minutes;
P = population;
S = average network speed, miles per hour; and
k_2 = a constant.

For two points in time (L_2 and L_1) and a constant population, the above expression can be stated as

$$\frac{L_2}{L_1} = \frac{S_2}{S_1}^{1.5} \tag{2}$$

The simulation study of the work trip indicates, however, that

$$\frac{L_2}{L_1} = \frac{S_2}{S_1}^{0.75} \tag{3}$$

Therefore, to derive a relationship between opportunity trip length and average network speed, three cases will be considered: Case A is for the 1.50 power, case B is for the 0.75 power, and case C is for the 1.00 power.

For case A, we know that average network speed can be defined as

$$S = \frac{L}{t} \tag{4}$$

Substituting equations (1) and (2) into equation (4) for two points in time will result in

$$\frac{S_2}{S_1} = \frac{S_2}{S_1}^{1.5} \frac{O_2}{O_1}^{-0.5}$$

Rearranging and solving for O_2/O_1 we obtain equation (5):

$$\frac{O_2}{O_1} = \frac{S_2}{S_1} \tag{5}$$

This indicates that the change in the mean opportunity trip length is directly proportional to the average network-speed change.

For case B, we substitute equations (1) and (3) into equation (4) for two points in time. This results in

$$\frac{S_2}{S_1} = \frac{S_2}{S_1}^{0.75} \frac{O_2}{O_1}^{-0.5}$$

Rearranging and solving for O_2/O_1, we obtain equation (6):

$$\frac{O_2}{O_1} = \frac{S_2}{S_1}^{-0.5} \tag{6}$$

This indicates under simulation study conditions that the change in the mean opportunity trip length would be inversely proportional to the square root of the speed change.

For case C, the relationship between opportunity trip-length change and the speed change is

$$\frac{S_2}{S_1} = \frac{S_2^{1.00}}{S_1} \frac{O_2^{-0.5}}{O_1} \tag{7}$$

$$\frac{O_2}{O_1} = 1.00$$

This indicates that there would be no change in the mean opportunity trip length with changes in network speed.

Change in trip length is caused not only by network speed, but also by changes in city structure as shown in Figure 2 for the city of Washington. The trip length in minutes increased over 14 percent during this period (Fig. 10). There was little concurrent change in the speed of the transportation network, and therefore the trip-length change was caused primarily by change in the city structure. Trip length is sensitive, therefore, to changes in speed of the transportation network and changes in the structure of the city.

NETWORK SPEED AND CITY STRUCTURE

As previously indicated, changes in the work-trip length in miles is proportional to the 0.75 power of the change in network speed. But the historical data available on trip length in five cities—Detroit, Baltimore, Washington, Sioux City, and Reading—indicated that the trip length in miles increased according to the 1.5 power of the change in network speed. The difference between the 0.75 power and the 1.5 power must reflect the amount of change that was a result of the change in the urban structure, where the cities tended to spread out. Whether this pattern will continue into the future is difficult to say. There seems to be evidence at the present time, however, that cities are increasing their residential densities and are not spreading out as much as they did just after World War II. The recent apartment house expansion is an illustration of this pattern.

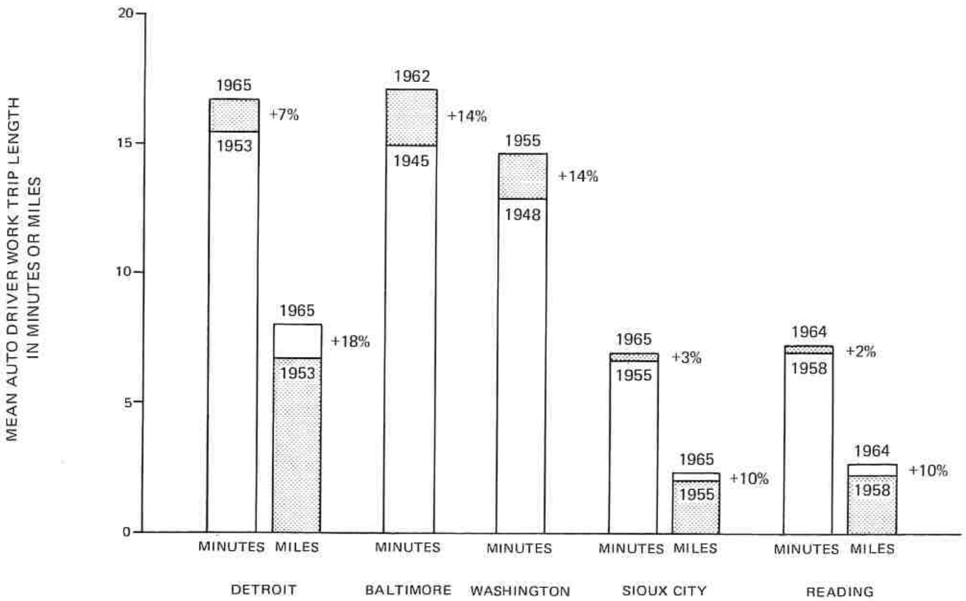


Figure 10. Historical increases in the work-trip length. (See Fig. 8 for source.)

TABLE 1

THE WORK-TRIP LENGTH BY GEOGRAPHIC AREA: DETROIT (1965)

Geographic Area	Percent of Total Auto Trips	Mean Trip Length[a] (min)	Percent Difference From Total/Average
CBD	2	29.5	+70
Central	33	20.1	+16
Other	65	15.9	-8
Total/average	100	17.3	

[a]Mean trip lengths do not include terminal times (6 minutes in the CBD and about 2 to 4 minutes in other areas).

The socioeconomic structure of the metropolitan area also has an effect on the work-trip length. In Detroit in 1965 it was found that the average work-trip length for certain geographic areas varied by as much as 70 percent from the metropolitan-wide average (Table 1). This large difference was caused by spatial separation of work-trip opportunities for different socioeconomic classes (Table 2). The data indicate that the actual trip length and the opportunity work-trip length increase as the income of the trip-maker increases. This explains the differences in trip length in relation to income that have been found in other metropolitan areas. The fact that the rich or poor have work-trip lengths that are not consistent from one metropolitan area to another is really a result of the differences in the spatial distribution of work-trip opportunities. Low-income workers in the CBD of Detroit find themselves surrounded by many job opportunities. The job opportunities relevant to their skill levels, however, are often found at some outlying manufacturing plant. The related average opportunity-trip as well as the actual work-trip length is therefore longer than the average for the metropolitan area.

Another way to explore whether this change in trip length is likely to continue in the future is to attempt to use the proposed relationships between urban travel and city structure to see what probably has occurred in the past as transportation systems have been improved and as cities have changed.

TRIP LENGTH AND CITY DEVELOPMENT

The kind of trip length that might be expected in a city can be calculated if certain assumptions about the average network speed are made and if it is assumed that the impact of the change in network speed is related to the 0.75 power, the 1.0 power, or the 1.5 power of the change in network speed. Figure 11 shows the work-trip length in terms of minutes for different network speed—3, 10, 20, and 30 mph—using the exponents of 0.75, 1.0, and 1.5. It was also assumed that the average work-trip length would be about 20 minutes, with network speed of 30 mph. (This average length is generally found in cities of approximately 2 to 3 million people.)

These results indicate that in a pedestrian society (3 mph) with a corresponding high-density city structure, the average work-trip time with today's automobiles would be down to about 3 to 5 minutes. Although historical data on the pedestrian society are not available, one does have indications from literature such as Dickens' work dealing with London that people did walk more like 20 minutes to work rather than 3 minutes. Such

TABLE 2

RELATIONSHIP BETWEEN THE MEAN WORK-TRIP LENGTH AND THE MEAN OPPORTUNITY-TRIP LENGTH: DETROIT (1965)

Income Class (dollars)	Actual Trip Length	Ratio of Actual to Average	Opportunity-Trip Length	Ratio of Opportunity-Trip Length to Average Trip Length
0-3,000	14.4	0.83	28.3	0.92
3,000-5,000	15.7	0.91	28.2	0.92
5,000-7,000	15.7	0.91	29.2	0.95
7,000-10,000	17.8	1.03	30.9	1.01
10,000 and over	19.6	1.14	31.1	1.02
Average	17.3		30.6	

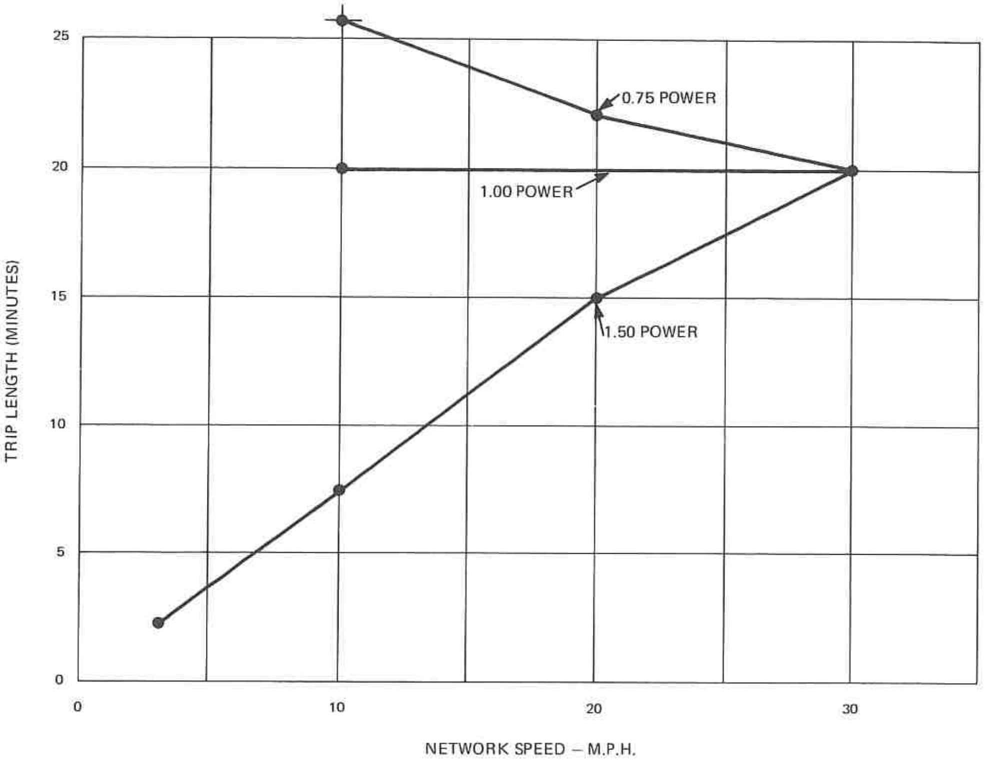


Figure 11. Relationship between network speed and trip length (minutes).

literary references seem to indicate that people are traveling about the same time to work as they did 50 to 100 years ago for comparable size cities. Thus, perhaps the impact that network-speed change has on trip-length change is more nearly related to the 1.0 power than to the 1.5 power.

CHANGES IN ACCESSIBILITY

A recent report of Bieber and Jorry (4) shows that the network speed for typical European cities is about one-half that for comparable U. S. cities and that the average distance from the CBD to the center of population in Europe was about one-half the average distance in the United States. This would indicate that the average opportunity distance in the United States and Europe for comparable size cities is nearly the same.

Figure 12 shows how the mean opportunity length changes with various exponents (0.75, 1.0, and 1.5) for changes in network speed. The results indicate that if the work-trip length goes up in proportion to the 1.5 power there is a substantial reduction in the average length of trip opportunities. This reduction would imply that cities are becoming less convenient in terms of time. On the other hand, if trip length goes up at the 1.0 power related to network speed, there would be no change in opportunity; this is probably what has been happening, as indicated in the comparison between European and United States cities. This relationship raises a very basic question. Are we providing a better transportation system if we neither decrease the average opportunity trip length nor hold it constant?

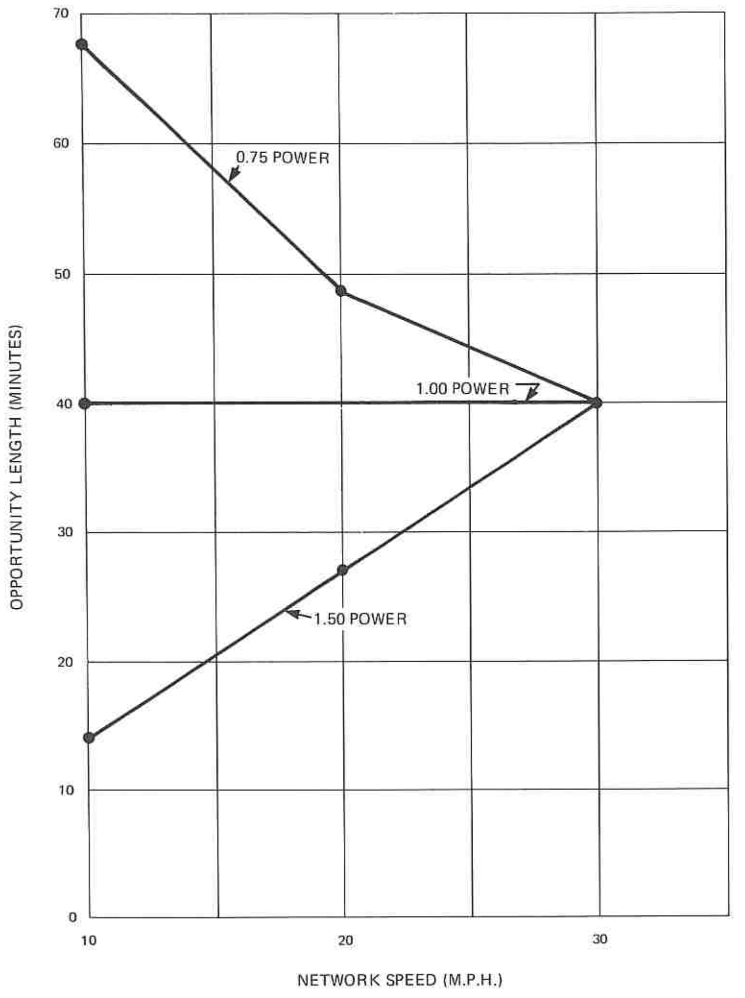


Figure 12. Relationship between network speed and opportunity length.

URBAN DEVELOPMENT AND TRANSPORTATION OBJECTIVES

Urban and transportation planners have allowed for lower density of development, both residential and commercial. Certainly there are strong indications that people prefer lower residential densities, but our residential areas could be developed to improve accessibility to jobs and other activities, particularly in the first 10 to 15 minutes of travel. Low-density commercial areas have allowed us to take advantage of new methods of manufacturing and handling of goods, but again these areas could have been structured and related to residential areas to improve accessibility.

What has become of transportation costs with the improvement of the transportation system? As shown in Figure 13, regardless of the assumption that we make on the value of the exponent, it is quite clear that the work-trip length in miles increases substantially with increases in network speed. For example, if we increase the average speed of a transportation network from 20 to 30 mph we could be increasing the work-trip length from 40 to 100 percent.

As has been demonstrated by the various needs studies performed by the Automotive Safety Foundation, Washington, D. C., the increases in transportation costs are related to the work-trip length and the amount of travel involved. Thus the average cost of

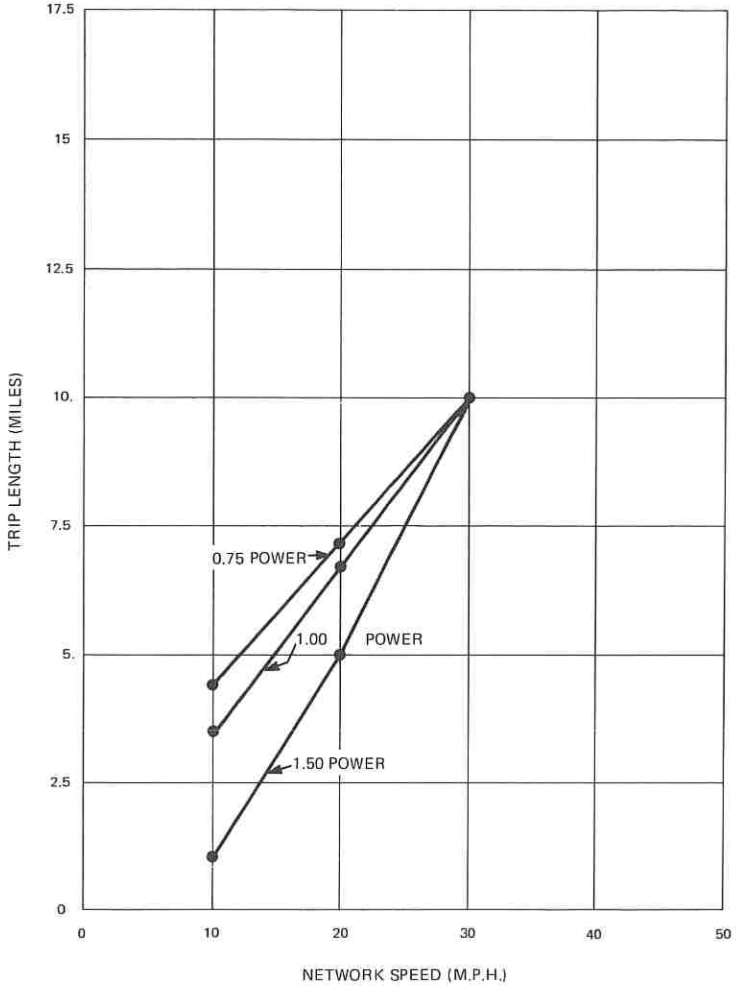


Figure 13. Relationship between network speed and trip length (miles).

providing transportation service has been going up with the improvements in transportation systems, resulting in lower residential density patterns.

SUMMARY

Urban travel as measured by the work-trip length (in miles and minutes) is highly related to city structure, which is reflected by the distribution of job opportunities within a metropolitan area.

Increases in the speed of transportation systems (a) cause the average work-trip length to increase, (b) increase the average length of job opportunities, and (c) allow for lower residential densities within the metropolitan area. Conversely, slower speeds result in lower work-trip lengths, reduced average lengths of job opportunities, and higher densities (urban living).

The selection of city structure and broader considerations relating to the environment and living preferences are the key decisions that must be made. Once city structure and environmental objectives are selected, care must be exercised by the planner to develop a transportation system that is directed towards that particular city structure and to assure that the broader environmental considerations are met.

REFERENCES

1. Mitchell, Robert B. Metropolitan Planning for Land Use and Transportation. The White House, Office of Public Works Planning, Washington, D. C., December 1959.
2. Origin and Destination: Methods and Evaluation—10 Reports. Highway Research Record 114, 1966.
3. Clark, Colin. Urban Population Densities. Jour. Royal Statistical Soc., Series A, Vol. 114, No. 4, 1951, pp. 490-498.
4. Bieber, A., and Jorry, J., under the direction of G. Saias. Formes Urbaines et Systems de Transport Dans Huit Métropoles de Moyenne Importance. Setec-Economie, 15, Quai Paul Doumer, 92 Courbevoie.

Action Space Formation: A Behavioral Approach to Predicting Urban Travel Behavior

FRANK E. HORTON and DAVID R. REYNOLDS,
Institute of Urban and Regional Research, University of Iowa

A conceptual framework for examining the process of action space formation, a process important in the understanding of urban travel behavior in urban areas, is delineated. Urban travel behavior is considered a direct product of the action space of households living within urban areas. Action space formation is considered the outcome of a two-stage process. The first centers on the relationships between the householders' residential location, length of residence, socioeconomic attributes, objective urban spatial structure, and travel preferences and perception of the urban environment. The second stage treats the interrelations of travel preferences, perception of the urban environment, and the action space formation process. An analytical framework, designed to test and calibrate the conceptual framework, is presented. Preliminary empirical findings related to the structure of aggregative action spaces of two quite different urban neighborhoods are presented. Differences in the structure of route preferences in a shopping context for the two neighborhoods are discussed. Linkages between the action space of an urban household and its urban travel behavior are defined in the empirical analyses.

•THIS PAPER relates the concept of action space formation to urban travel behavior and describes several empirical analyses considered useful in constructing a behavioral model of action space formation, which has as one of its outputs the specification of urban travel behavior. Hence, the concept of action space seems to be extremely useful when applied to the problem of understanding and predicting urban travel patterns.

Before proceeding, we shall define several basic concepts used throughout the paper. The term "objective spatial structure" refers to the location of a household relative to the actual locations of all potential activities and their associated objective levels of attractiveness within an urban area. "Action space" is the collection of all urban locations about which the individual has information and is the subjective utility or preference he associates with these locations. The subjective utility or preference is evaluated with regard to both potential and actual travel behavior. Geometrically, action space is characterized by two components: first, its spatial extent as defined by the set of locations; and, second, a generalized surface (both piecewise and continuous) specifying the utility or preference level associated with each location. This definition is a slight reformulation of the concept as originally presented by Wolpert (1). An individual's "activity space" is defined as the subset of all urban locations with which the individual has direct contact as the result of day-to-day activities. The term is similar to the notion of activity system put forth by Chapin (2) and Hemmens (3). Geometrically, activity space is characterized as a surface (again both piecewise and continuous) descriptive of the intensity of actual travel behavior over portions of the action space.

Although the individual theoretically has access to a broad range of environmental information spanning local to international levels, usually only a limited portion of the urban environment is relevant to this travel behavior in any given context. Even though

Paper sponsored by Committee on Origin and Destination.

the individual's action space is limited spatially, a meaningful examination of its formation is likely to include consideration of a wide range of travel behavior (e.g., the journey to work, shopping, visiting neighbors, etc.) and cannot ignore the individual's perception of the objective spatial structure of his physical, economic, and social environment within which this behavior takes place.

The degree to which the individual's travel behavior is in equilibrium with the objective spatial structure of the city depends upon his ability to collect and assimilate information concerning it. Differences in race, sex, education, income, and social status are all likely to represent early and salient factors contributing to the individual's efficiency in receiving and weighing such information and may also induce distinctive biases in his travel behavior. For example, urban ghetto life would seem to generate a cramped view of the urban environment partly because of the spatial concentration of environmental experiences (4).

Because no two individuals perceive the city from exactly the same point simultaneously and because each bases his interpretation of information gleaned from the urban environment on past experience, action spaces vary from person to person. Although perceptions and action spaces are to a degree individualistic, there is reason to suggest that they are shared, to a large extent, by like groups of people. For example, the formation of the individual's action space and its manifestation in urban travel behavior is almost certainly affected by his group memberships, his position in social networks, his position on one of his divergent life cycles, and his spatial location with respect to potential trip destinations in the environment. Obviously, the latter (generally his residence) is the primary node in any action space, as it is in any study of urban travel behavior. Through personal observation, the individual is likely to be more familiar with local areas (the areas in the vicinity of his residence and his workplace, in particular) than those points at greater distances from him and about which available information is limited. Thus, a person's environmental perception is conditioned only in part by the nature of the objective milieu itself.

Isard (5) has suggested that variations in individual space and time preferences are so great as to preclude any economic rationalization of individual travel behavior. Thus far, Isard's pessimism seems to have been justified in that deterministic economic models, with their built-in assumptions of economic rationality, have been noteworthy for their lack of success in accounting for spatial behavior, except at a highly aggregative level. It would appear appropriate, therefore, to adopt a behavioral approach that examines the formation of the individual's action space and his resulting travel behavior as a function of his socioeconomic characteristics, his cognitive images of the urban environment, and his preferences for travel.

The individual's perception cannot be viewed as being static, but rather as changing via a complex learning process. Given socioeconomic constraints, and provided that the individual does not change his place of residence, one would expect that as an individual modifies his perception of his environment by traveling within it and by communicating with his peers about it, his travel behavior would approach spatial equilibrium and "rational economic behavior." To expect such behavior, however, would also presuppose that the objective spatial structure itself does not change. In fact, the objective spatial structure and its components (i.e., retail structure, location of employment opportunities, temporal differences, residential quality, etc.) are constantly undergoing change that results in the continuous reordering of the perceived urban spatial structure.

Technological change also plays an important role in extending an individual's action space and in modifying its morphology; but, at the same time, such change impedes the tendency of the individual's travel behavior toward the economically rational. An appropriate example is the increase in the use of private transportation, which has enabled the individual to extend the perimeter of his action space. At the same time, however, continued increase in the use of this mode of transportation has affected spatial distortions and place disutilities within the areally increased action space: the penalties are congestion with all its concomitant psychological and physiological effects. Therefore, continuous technological change and perceptual lags would seem to prevent the individual from achieving spatial equilibrium and economically rational behavior, although he may approach these states. Nevertheless, an individual's perception or cognitive image of

the city should be in equilibrium with his action space, the one being a satisfactory predictor of the other.

RELEVANCE OF ACTION SPACE FORMATION RESEARCH TO STUDIES OF URBAN TRAVEL BEHAVIOR

The nature of urban travel patterns and the factors that condition them have prompted numerous research endeavors. An examination of these travel studies reveals a divergence in the factors that are reported to be of greatest value in forecasting travel patterns. Although this divergence may be the result of the varying purposes for which these studies were conducted and of the aggregate data upon which most are based, the factors nonetheless are inadequate for providing a clearly articulated empirical basis for the development of a theory of urban travel behavior. Attempts to use variables that have been shown to be related statistically to urban travel at the aggregate level have been almost universally unsuccessful at the household level. This failure is not surprising; factors that are important conditioners of mass group behavior (such as employment rate, median income, etc.) are devoid of behavioral meaning at a less aggregated level.

Urban activities provide the impetus for movement. An individual's perception of their desirability and location form the basic psychospatial milieu within which action space formation takes place. Thus, the location and spatial structure of urban activities are important to the modeling and understanding of action spaces. Clearly, a better understanding of the urban household's travel behavior demands that more research be directed toward discerning fundamental processes underlying this behavior. Action space formation is one such process.

Research concerned with action space formation would be an important initial step in ascertaining the impact of modifications in the location of activities and facilities in cities (e.g., urban renewal or expressway construction) upon the urbanite's cognitive image of, and travel behavior within, the modified urban system. For example, what effect does the construction of an expressway through a low-income, central-city area have upon the action spaces of those residents in the area relative to its impact on the action spaces of suburban commuters? What modifications of the urban environment maximize the spatial extent of action spaces summed over the entire metropolitan population? Research directed toward providing a framework on the basis of how such questions can better be posed and answered is of fundamental importance to both the social theoretician and the urban transportation planner.

Very few behavioral scientists have attempted to investigate systematically the formation of individual action spaces relative to various types of behavior. Wolpert's investigations (1) into the relationship between action space and the decision to migrate would seem to provide a useful framework for research into other areas of spatial behavior. However, the work of other social scientists—sociologists and social psychologists, in particular—also contains several useful insights into the formation of the individual action space and the extent to which these are shared by groups of people (6). The importance of social groups in affecting an individual's behavior has been studied especially well (7). Studies of the networks of interpersonal contacts have also provided empirical information which suggests a close relationship between social and spatial propinquity (7, 8). The flow of information through the networks of interpersonal communication is the basis for the images shared by individuals in social networks. Although group membership fosters the development of shared images, it may also inhibit the flow of information to the individual from other sources. In this regard, Deutsch (9) has commented on the way people may be "marked off from each other by communicative barriers, by 'marked gaps' in the efficiency of communication."

Some investigations, such as Lynch's empirical examinations (10) of how people in several areas perceive the spatial structure of their home cities, are highly suggestive of research hypotheses. The present state of knowledge is meager, however, concerning how the actual location of activities and the locational relationships between them, an individual's cognitive image of these activities and their locations, and travel preferences interact to form the action space.

CURRENT RESEARCH

There are several ways in which one could examine household action spaces and isolate those factors that are important inputs to their formation and change. The approach selected here is, at the outset, decidedly empirical because it involves collecting a considerable amount of information at the household level and testing sets of interrelated hypotheses. The data collection and hypotheses testing, however, is governed by the conceptual framework shown by Figure 1.

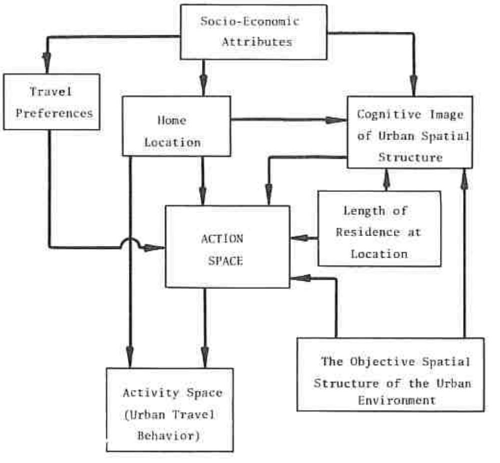


Figure 1. Conceptual model for action space.

The research currently being conducted consists of three closely related tasks:

1. The elucidation and testing of a behavioral model whose inputs interact to form an urban resident's action space and determine its spatial structure and extent, thereby defining the outer limits of likely travel within the urban area.
2. The definition and measurement of the morphology of action space and of cognitive images of the city so as to facilitate testing of theoretical hypotheses.
3. The development of methods for evaluating the effects of a changing objective spatial structure on the householder's environmental perception and his preferences for various urban travel activities.

Methodological Considerations

Two compact residential areas in Cedar Rapids, Iowa, constitute the study areas. These were selected with two considerations in mind: first, it was deemed important to proceed with interviews both in an area where the ratio of external to internal trips was high and in one where this ratio was low; and, second, it was necessary to contrast two populations from areas of quite different, but reasonably internally homogeneous, socioeconomic characteristics. The data that formed the basis of the study were obtained from the 1965 Origin-Destination Survey of the Cedar Rapids metropolitan area and from city census information. The areas selected were a low-income area in the central portion of the city, known locally as the Oak Hill-Jackson area, and an upper-middle-income area on the western perimeter of the city, known locally as Cedar Hills.

The interviewing phase of the investigation uses a pretested questionnaire designed to elicit the following information from approximately 200 sampled households having access to private transportation in each of the two areas:

1. Socioeconomic characteristics and group memberships;
2. Relative locational attitudes including past and present job location and locations of friends and relatives in Cedar Rapids;
3. Habitual routes of travel for the journey to work and for certain types of shopping;
4. Preferences for characteristics of travel, destinations, routes, and general trip structure;
5. The spatial extent and structure of the household's action space;
6. The perceived range of spatial choice in selected shopping situations; and
7. Perceptions of residential quality and shopping facilities in subareas of the metropolitan area.

This investigation is not concerned with the householder's composite image of Cedar Rapids. (Other studies, such as Lynch's (10), have discussed this image.) Rather, the present study is concerned with the way individuals structure their preferences for

spatial interaction in varying travel and shopping contexts (e.g., shopping trips, etc.). The interest in cognitive images of the environment focuses on the individual's spatial ordering of Cedar Rapids in terms of residential quality and shopping opportunities in general, as well as on his perception of individual retail establishments in terms of locational and qualitative attributes. This approach is taken because preferences and perceptions are used in an analytical rather than a descriptive manner—analytical, in the sense that this information will be input into a model of action space formation.

One problem that became immediately apparent was the operational articulation of the concept of action space. Because it was desirable for the concept to reflect not only the actual travel behavior of the individual, but also potential travel behavior, action space was operationally defined to be the area with which the individual perceives himself to be familiar. The structure of an individual's action space is defined by the correlation between areas of varying levels of perceived familiarity. To derive these measures, each sampled individual was confronted with a map dividing the Cedar Rapids metropolitan area into 27 subareas; and he was asked to indicate, on a five-point scale, the level of his familiarity with each subarea. These ordinal familiarity responses for each of the two samples were then transformed to an interval scale by a psychological scaling technique described in the following section. It was then possible to determine the basic structural dimensions of the aggregate action spaces of both sampled populations by employing "latent structure" or factor analytic procedures. The vectors of factor scores on each of the basic dimensions specify the location and structure of each individual's action space relative to the aggregate action space of his residential group.

Given the derived vectors depicting action spaces, it is necessary to operationally define the components related to action space formation shown by Figure 1. Socioeconomic attributes, home location, and length of residence for each household are obtained directly. Measures of the "objective spatial structure of the urban environment" are obtained from exogenous data sources such as Polk's Directory, 1966; 1968 school census materials; and several inventories developed in the 1965 Cedar Rapids Transportation Study. "Travel preferences" and "cognitive image of urban spatial structure" are both vectors of derived measures and, hence, require some elaboration.

The individual's travel preferences are derived from three sets of responses to questions in which the individual indicates the level of importance or preference for selected destination, travel, and route characteristics in varying travel contexts. Travel preferences include consideration of both time and distance. Responses were transformed into interval scale measurements and were analyzed as to levels of dimensionality by employing factor analytic procedures.

The measurement of cognitive image of urban spatial structure is on two levels. On the first, the individual evaluates his actual shopping destinations with respect to a given set of attributes. Because several of the attributes relate to the location of a destination, it is possible to ascertain the extent to which subjective destination characteristics distort objective travel distances and locational settings. On the second level, the individual evaluates residential quality and shopping facilities in each subarea of his action space. Evaluations are constrained to a five-point scale of "goodness" and are analyzed in the manner suggested.

PRELIMINARY EMPIRICAL FINDINGS

Before full-scale model construction can begin, action space and the salient inputs to its formation must be identified quantitatively. The empirical research deals with the aggregate structures of the action spaces of the two sampled populations and the aggregate structures of their preferences for types of routes in a multiple-purpose shopping context.

Scaling Procedures

The majority of the information was elicited from respondents in the form of a constrained-response questionnaire allowing the respondent to subjectively rate areas and variables on a categorical scale of 0 to 4. For example, an individual's familiarity with an area would range between 0 (unfamiliar) and 4 (very familiar). Given the need

for interval-scaled data in the majority of the analyses and in the actual model-building effort, it was necessary to transform the ordinal response data. This was accomplished by the application of a multidimensional scaling technique derived from Thurstone's law of categorical judgement. The methodology of this study is an adaptation of the work of Torgeson (11) and relies heavily on the modifications made by Peterson (12). Computer programs used in this analysis are variations of those written by Wachs (13).

In using the method, it is assumed that there is a psychological continuum that defines for a respondent the attribute he is considering. An attribute in question may be "importance," "goodness," "quality," "familiarity," etc. When the respondent considers a variable, a discriminal process enables him to place his image of that variable at a point on the psychological continuum corresponding to the perceived quantity of the attribute in question. The discriminative process is assumped to be probabilistic. Therefore, the respondent's placement of a value along the continuum will not be the same on every trial of an experiment, but instead will be defined by a Gaussian distribution on the continuum. The true value of the subject's response is taken to correspond to the mean of the Gaussian distribution of his responses. Each variable rated will have a particular mean response and dispersion about the mean.

For the purpose of this analysis, it is also assumed that the psychological continuum can be partitioned into a chosen number of ordered categories. If follows that a given category boundary will not always be perceived at a unique point on the continuum. Its location is also assumed to be defined by a Gaussian probability distribution with each category boundary having its own mean and dispersion. The operational problem, then, is to estimate the locations of each of the category boundaries and to assign each ordinal response to the midpoint location of that category on a psychological continuum, and to thereby convert the ordinal responses to interval-scaled values. As Torgeson (11) points out, this can be accomplished if one can assume that the location of any given category boundary is independent of the variables under consideration, and that the variability of the location of any given category boundary is constant across variables. These assumptions are met whenever the respondent maintains the same psychological continuum for the attribute under consideration over all variables. This does not appear to be unrealistic.

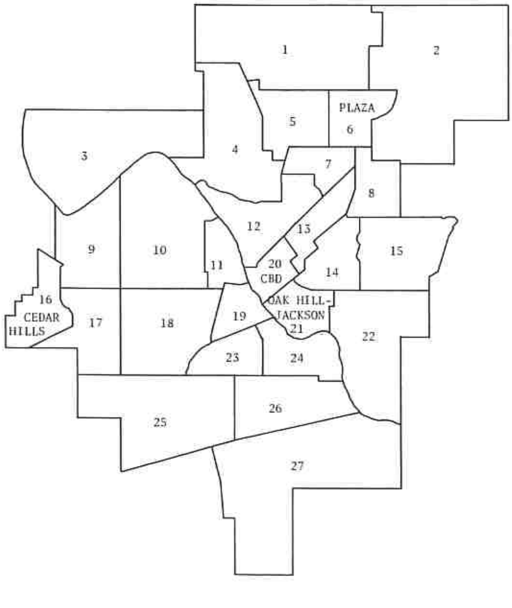


Figure 2. Subareas of metropolitan Cedar Rapids evaluated by respondents.

Aggregative Action Spaces for the Two Samples

As noted previously, the action space of an individual is operationally defined as the area with which he perceives himself familiar. (For the location of each of the 27 subareas evaluated, see Fig. 2.) Hence, it includes those areas within which the majority of his spatial behavior takes place as well as those that encompass potential interaction areas. Because the level of an individual's familiarity varies from one subarea to another, the areal structure of a composite action space for a residential community, such as those in this study, can be defined realistically by the correlations between areas of varying levels of perceived familiarity. If action spaces are highly individualistic and do not exhibit systematic regularities, then such correlations will be low and the concept of action space will be of dubious generality in modeling urban travel behavior. Therefore, the first task was to examine the common dimensions of variability in the spatial structures of the action spaces for each of the two sampled populations.

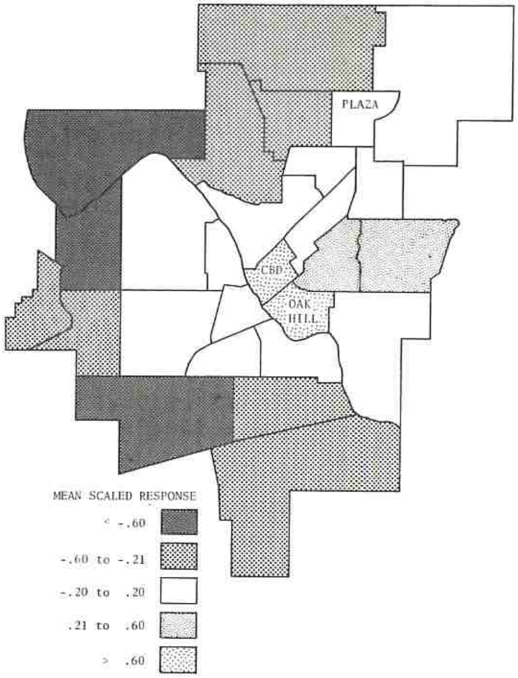


Figure 3. Familiarity-mean scaled responses—Oak Hill-Jackson.

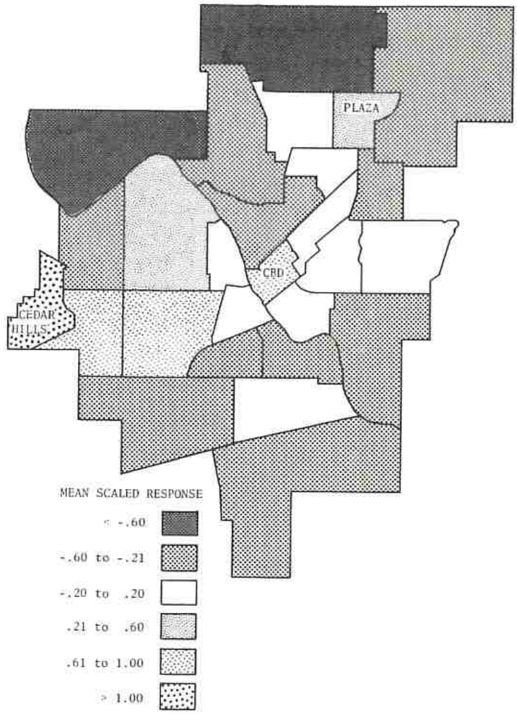


Figure 4. Familiarity-mean scaled responses—Cedar Hills.

The ordinal familiarity responses for the 27 subareas of metropolitan Cedar Rapids were transformed to an interval scale for each of the two samples. The mean scaled familiarity responses for the sample in the low-income, central-city residential area, Oak Hill-Jackson, are shown in Figure 3. As can be seen, the aggregative action space of this group appears, in general, to be characterized by a pronounced familiarity with the home area and the adjacent central business district (CBD) and by an increasing unfamiliarity with increasing distance from the home area. A major exception to this generalization is a high level of familiarity with areas comprising the Northeast Corridor, which is traversed by the most important thoroughfare in the city. Clearly, levels of familiarity are related to urban spatial structure as perceived by the respondents.

Similarly, Figure 4 shows the mean scaled familiarity responses of the sample from the middle-income residential area, Cedar Hills. The aggregate action space of this residential group differs from that of the sample from the low-income, central-city area, primarily in the more pronounced overall familiarity and in the linear rather than modal pattern of familiarity levels. Areas of highest familiarity are the home area and those in the direction of the CBD, the CBD itself, and that area containing the major outlying shopping plaza. In general, familiarity decreases with distance from the home-CBD-shopping-plaza axis of familiarity. It should be noted that the same general elements of urban spatial structure were evaluated similarly by both sets of respondents. Because of the location of the home in relation to those elements, however, the pattern of mean scaled familiarity was circular in the case of Oak Hill-Jackson and linear in the case of Cedar Hills.

To assess the extent to which there are common dimensions of variability in the spatial structures of the composite action spaces for each of the two sampled populations and to identify the spatial dimensions common in each, the matrices of correlations

representing familiarity between areas for each sample were subjected to principle components analyses with subsequent varimax rotation. In the resulting factor structures, each of the 27 areas had a high factor loading on one and usually only one of the basic spatial dimensions of variability in familiarity extracted (see Tables 1 and 2). Therefore, the variation accounted for by each basic dimension was composed primarily of systematic covariations in familiarity within clusters of areas.

For the Oak Hill-Jackson sample, six dimensions of variability in familiarity which accounted for 67 percent of the total variance in between-area familiarity were extracted (see Table 1). These are shown in Figure 5. The areal partitioning of the metropolitan area accomplished through this analysis resulted in contiguous and nonoverlapping groupings of subareas. The resulting spatial dimensions of variation in familiarity in terms of decreasing levels of "explained" variation can be characterized as follows:

1. An area of uniformly moderate familiarity;
2. The home area and maximum familiarity;
3. A northern area of uniform unfamiliarity;
4. A western area of uniform unfamiliarity;
5. An area of decreasing familiarity with increasing distance to the south of the home area; and
6. A concave surface of familiarity peaking in the CBD and the major outlying shopping plaza.

In the Cedar Hills sample, the principle components analysis resulted in the extraction of seven dimensions for the aggregate action space (Fig. 6). These dimensions account for 69 percent of the between-area levels of familiarity (see Table 2). With only one exception, the areas comprising these basic dimensions are contiguous and

TABLE 1

AGGREGATE STRUCTURE OF THE OAK HILL-JACKSON ACTION SPACE RESULTING FROM THE VARIMAX FACTOR STRUCTURE

Subarea[a]	Loadings on Each Factor[b]					
	1	2	3	4	5	6
1			0.644			
2			0.542			
3			0.766			
4			0.631			
5	0.367		0.692			
6						0.603
7			0.450			0.560
8			0.438			0.530
9				0.710		
10	0.760					
11	0.770					
12	0.447					
13	0.408					0.482
14						0.710
15					0.361	0.703
16				0.723		
17				0.749		
18	0.606			0.392		
19	0.456	0.354				
20						0.731
21		0.842				
22		0.544			0.505	
23	0.409				0.568	
24					0.573	
25				0.439	0.623	
26					0.688	
27				0.369	0.657	
Percent of variance	43.1	6.6	6.3	4.0	3.9	3.3

[a]For the location of subareas, see Figure 2. [b]Only factor loadings greater than 0.35 are shown.

TABLE 2

AGGREGATE STRUCTURE OF THE CEDAR HILLS ACTION SPACE RESULTING FROM THE VARIMAX FACTOR STRUCTURE

Subarea[a]	Loadings on Each Factor[b]						
	1	2	3	4	5	6	7
1					-0.766		
2					-0.809		
3		0.765					
4		0.486					
5				-0.436	-0.503		
6				-0.854			
7				-0.712			
8						-0.476	0.407
9						-0.708	
10						-0.646	
11						-0.571	
12					-0.385	-0.472	
13				-0.562			
14							0.767
15							0.739
16			0.831				
17			0.657				
18	0.433		0.492			-0.366	
19	0.657						
20	0.571			-0.409			
21	0.574					-0.373	0.361
22	0.432						0.579
23	0.763						
24	0.774						
25	0.647					-0.400	
26	0.644						
27	0.665						
Percent of variance	40.7	7.2	5.1	4.5	4.1	3.6	3.4

[a]For the location of subareas, see Figure 2. [b]Only factor loadings greater than ±0.35 are shown

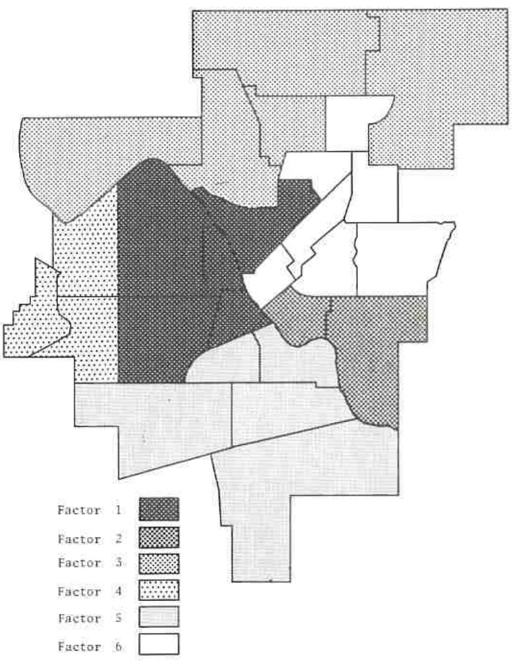


Figure 5. Familiarity factor structure—Oak Hill-Jackson.

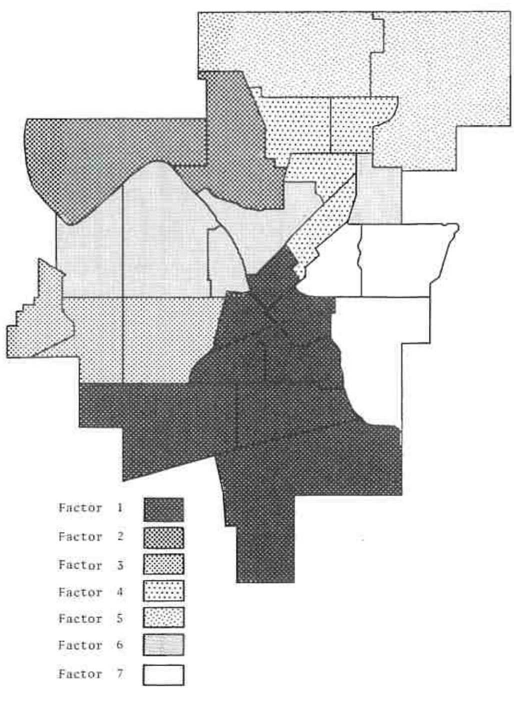


Figure 6. Familiarity factor structure—Cedar Hills.

nonoverlapping. The major dimension of variations in between-area familiarity is accounted for by variations within area 1, which is characterized by decreasing familiarity from the CBD in a southern direction. The remaining dimensions of variation in decreasing order of explained variation may be characterized as follows:

Area 2, a relatively inaccessible, low population density area of very low levels of familiarity.

Area 3, an area of decreasing familiarity out from the home area and towards the CBD.

Area 4, an area along the major thoroughfare axis characterized by increasing familiarity that peaks at the major shopping plaza.

Area 5, a northeast area of uniform unfamiliarity.

Area 6, an area of moderate familiarity to the northwest of the major familiarity axis.

Area 7, an area of moderate familiarity to the southeast of the major familiarity axis.

TABLE 3

SHOPPING TRIP ROUTE-SELECTION ATTITUDE FACTORS AND FACTOR LOADINGS—CEDAR HILLS

Factor Number and Name	Percent of Variance	Variable Name	Loading
1. Preference for less traffic	31.22	Fewer trucks and buses	+0.694
		Less automobile traffic	+0.670
		Shopping can be completed without making left-hand turns across traffic	+0.625
		Few pedestrians crossing	+0.559
		Route is safer than others	+0.535
2. Preference for convenient nonredundant route and shopping node geometry	7.52	Shopping can be completed without any doubling back or retracing of original route of travel	+0.693
		On this route it is possible to stop at more than one shopping center rather than traveling downtown	+0.676
		Most of the stores are near home on this route	+0.649
		Stores are only a short distance apart on this route	+0.608
		On this route it is possible to travel to most distant store and then do the rest of my shopping on the way home	+0.527
		It is possible to return home from the last stop in the shortest time possible	+0.492
3. Preference for steady higher speed	5.38	Pavement is smoother	+0.801
		Possible to maintain a steadier rate of speed	+0.599
		The speed limit is higher	+0.534
4. Preference for less travel time	4.91	Trip takes less time	-0.770
		Fewer stops and interruptions	-0.689
		Distance is shorter	-0.683
		Route is more direct	-0.653
		Fewer traffic lights	-0.515
		It is possible to return home from the last stop in the shortest time possible	-0.510
		The speed limit is higher	-0.483
		It costs less to drive on this route	-0.459
5. Preference for pleasant scenery	4.48	Route goes through more pleasant neighborhoods	+0.861
		The scenery is more pleasant	+0.808
6. Preference for change in scenery	3.67	On this route it is possible to return home by taking another which takes no more driving time	-0.741
		This route goes through as many different parts of the city as possible	-0.671
		On this route it is always easy to determine how far I am from my home and destination	-0.461
7. Preference for safe, major arterials	3.43	There are more lanes on this route	-0.765
		The lanes are wider on this route	-0.717
		On this route it is easy to turn off onto alternative routes, if necessary	-0.521
8. Preference for clearly marked and easily followed routes	3.27	Route is easy to follow and remember	-0.692
		Route is more clearly marked	-0.613
9. Preference for auto services along route	3.16	There are more service stations along this route	-0.774

The unexpectedly high levels of explained variation resulting from the components analyses lend further credence to the notion that individuals sharing similar residential locations in an urban area share similar images of urban spatial structure and share similar action spaces. The analyses of the responses for both the Oak Hill-Jackson and Cedar Hills samples resulted in contiguous areas representing independent dimensions of familiarity. Thus, there are probably significant subgroups in each sample with remarkably similar action spaces in terms of both spatial extent and structure. Such subgroups are hypothesized to be characterized by similarities in travel and shopping preference structures, lengths of residence, location of work place, and ethnic and socioeconomic characteristics.

That different dimensions are characterized by similar levels of familiarity is indicative of significant subgroup differences most likely based on length of residence within the Cedar Rapids metropolitan area. This result may indicate a bias generated by stages in a learning process. The learning-process influence is also suggested by the apparent arterial bias exhibited in the familiarity levels of the Cedar Hills sample population. Because Cedar Hills is a newly developed subdivision in Cedar Rapids, the sample was primarily composed of residents with only short term experiences in the

TABLE 4

SHOPPING TRIP ROUTE-SELECTION ATTITUDE FACTORS AND FACTOR LOADINGS—OAK HILL-JACKSON

Factor Number and Name	Percent of Variance	Variable Name	Loading
1. Preference for safe major arterials	27.58	Route is safer than others	+0.749
		There are more lanes on this route	+0.725
		The lanes are wider on this route	+0.690
		It costs less to drive on this route	+0.503
2. Preference for less travel time	7.63	Distance is shorter	+0.845
		Trip takes less time	+0.741
3. Preference for convenient nonredundant route and shopping node geometry	5.64	Stores are only a short distance apart on this route	+0.782
		Shopping can be completed without any doubling back or retracing of original route of travel	+0.631
		Most of the stores are near home on this route	+0.626
4. Preference for pleasant scenery	5.50	The scenery is more pleasant	-0.867
		Route goes through more pleasant neighborhoods	-0.786
		This route goes through as many different parts of the city as possible	-0.436
		On this route it is possible to stop at more than one shopping center rather than traveling downtown	-0.434
5. Preference for well-known routes with auto services available	4.85	On this route it is always easy to determine how far I am from home and from my destination	+0.704
		There are more service stations along this route	+0.626
		On this route it is possible to return home by taking another which takes no more driving time	+0.519
6. Preference for less congestion and higher speed	4.48	The speed limit is higher	+0.646
		Shopping can be completed without making left-hand turns across traffic	+0.590
		Fewer pedestrians crossing	+0.502
		Less automobile traffic	+0.503
7. Preference for clearly marked routes	4.18	Route is more clearly marked	+0.680
		On this route it is possible to travel to most distant store and then do the rest of my shopping on the way home	+0.617
		Route is easy to follow and remember	+0.513
		On this route it is easier to turn off onto alternative routes if necessary	+0.475
8. Preference for steady speed	3.60	Easier to see what is ahead along this route	+0.698
		Possible to maintain a steadier rate of speed	+0.680
		Pavement is smoother	+0.672
		Route is more direct	+0.659
		On this route it is easier to turn off onto alternative routes if necessary	+0.449
9. Preference for less traffic	3.20	Fewer trucks and buses	+0.750
		Fewer stops and interruptions	+0.591
		Less automobile traffic	+0.517
		Fewer traffic lights	+0.468
		Fewer pedestrians crossing	+0.456

metropolitan area. This arterial bias, of course, is also indicative of the marked influence of the transportation network and, hence, the objective spatial structure of urban environments on action space formation.

Attitude Toward the Selection of Routes to Shopping Destinations

The analysis of the preferences for route selection in a multiple-purpose shopping context for both the Cedar Hills and Oak Hill-Jackson area proceeded in a manner similar to that outlined previously. The resulting factor structures derived from the route preference analysis are given by Tables 3 and 4. Nine factors accounted for 67 percent of the variance in both Cedar Hills and Oak Hill-Jackson.

Several of the factors in each of the tables are similar (see Tables 3 and 4). It is interesting, however, to note that the factor accounting for the largest amount of variance in the analysis of the responses made by the residents of the central-city community is the preference for safe, major arterials. This factor is indicative of a locational bias generated by congested travel conditions in their home area that make travel seem difficult in any direction. The factor accounting for the most variance in the Cedar Hills analysis indicates a preference for less traffic. This factor seems to indicate that Cedar Hills residents must drive into or through the CBD in many shopping situations. A preference for less travel time by the Oak Hill-Jackson respondents is also indicative of congested conditions in the sense that these people live much closer to a majority of the shopping opportunities yet feel that it takes more time to reach them. Cedar Hills respondents, on the other hand, reflect a need for routes that will lead them to more shopping opportunities closer to their home location and that would allow them to by-pass or avoid the downtown congestion indicated by the second factor.

CONCLUSIONS

The implications of these analyses for understanding the shape and extent of action spaces and the derivation of travel behavior characteristics are varied. In general, it may be concluded that the respondents in the Cedar Hills area will select routes that minimize congestion and allow for multiple-purpose trips. It would appear that multiple-purpose trips are more frequent for Cedar Hills residents because their location away from shopping opportunities forces them to evaluate the necessity of each trip. The Oak Hill-Jackson residents, on the other hand, will select routes that are major arterials and that decrease the amount of time necessary to complete their activities. The urban transportation system in Cedar Rapids has developed in such a way that north-south travel is impeded and east-west travel is facilitated. The location of the Oak Hill-Jackson area is such that residents must travel to the north to get either to the CBD or the major shopping center. Further, the Cedar Hills respondents are located at the western edge of the major east-west arterial and thus must continuously use this route to travel either to the CBD or beyond the CBD. As transportation system improvements and changes in the shopping opportunity configuration occur, one may conclude that the Cedar Hills activity space will decrease according to their preferences, whereas the activity space of the Oak Hill-Jackson residents will increase.

The empirical findings presented form the basis for several components of the conceptual framework outlined in Figure 1. Still to be completed are analyses to assess the variability in perception attributable to differences in socioeconomic attributes, length of residence, work location, and the spatial structure of friendship networks and organizational memberships. Definition of subgroups with like action spaces and travel behavior within the two sampled areas currently are being completed.

Probability matrices will be developed that are descriptive of changes in urban environmental perception with varying length of residence in the area. These matrices will be predicated on a detailed cross section of the sample, after controlling for nonhomogeneous socioeconomic characteristics. Further analyses will be designed to assess the variability in action space structure related to travel preferences and perception differences. It is envisioned that the final structure of the model will be defined by a set of simultaneous linear equations in which all components and each action space measure are included. The exact form of the equation set will be determined by the

outcome of the analyses previously described. Specification of the parameters in the model will allow for considerable experimentation as to the effects of changes in the objective urban environment of travel behavior. Further, the behavioral model of urban action spaces and its extension to travel behavior will provide an improved basis for predictive models in the urban transportation planning process.

ACKNOWLEDGMENT

This paper is based on research supported by The National Science Foundation.

REFERENCES

1. Wolpert, J. Behavioral Aspects of the Decision to Migrate. Papers and Proc. of the Regional Science Assn., Vol. 15, 1965, p. 163.
2. Chapin, F. Stuart, Jr. Activity Systems and Urban Structure: A Working Schema. Jour. Amer. Inst. Planners, Vol. 34, No. 1, 1968, pp. 11-18.
3. Hemmens, George C. Analysis and Simulation of Urban Activity Patterns. Paper Presented at the Annual Symposium on Application of Computers to the Problems of Urban Society, New York, Oct. 1968.
4. Suttles, G. D. The Social Order of the Slum. Univ. of Chicago Press, Chicago, 1968.
5. Isard, W. Location and Space Economy. M.I.T. Press, Cambridge, 1965, pp. 84-5.
6. Rubenstein, A. H., and Haberstrol, C. J. Some Theories of Organization. Richard D. Irwin, Inc., Homewood, Ill., 1964.
7. Festinger, L., Schacter, S., and Back, K. Social Pressures in Informal Groups. Harper and Brothers, New York, 1950.
8. Merton, R. K. The Social Psychology of Housing. In Current Trends in Social Psychology. Univ. of Pittsburgh Press, Pittsburgh, 1951.
9. Deutsch, K. M. Nationalism and Social Communication. M.I.T. Press, Cambridge, 1953.
10. Lynch, K. The Image of the City. M.I.T. Press, Cambridge, 1960.
11. Torgeson, W. S. Theory and Method of Scaling. John Wiley and Sons, New York, 1958.
12. Peterson, G. L. Subjective Measures of Housing Quality: An Investigation of Problems of Codification of Subjective Values for Urban Analysis. Northwestern Univ., Evanston, Illinois, PhD dissertation, 1965.
13. Wachs, M. Evaluation of Engineering Projects Using Perceptions of and Preferences for Project Characteristics. Research Report, The Transportation Center, Northwestern Univ., Evanston, Illinois, 1967.

Micro-Assignment: A New Tool for Small-Area Planning

GEOFFREY J. H. BROWN and ROBERT S. SCOTT, Creighton, Hamburg, Inc.

•EARLY in the past decade, the main impetus for the development of traffic assignment techniques was the need to prepare link volume estimates for region-wide networks, with emphasis on the accurate estimation of expressway volumes.

Around the middle of the decade, the first operational techniques appeared that dealt simultaneously with mixed transit/highway networks; and in at least one large metropolitan region (a region much too vast to be digested by existing assignment methods), a technique for directly estimating link volumes was developed that made it unnecessary to treat the entire region as a single indivisible entity.

In parallel with the growth in sophistication of assignment methods, an enormous increase has occurred in the last 10 years in both the number and quality of techniques for studying the behavior of vehicles at what might be called the microscopic scale—wending their way through intersections, forming queues at traffic lights and toll booths, or merging onto expressways. The models that have arisen from these studies have proven invaluable in the solution of problems ranging in complexity from fixing the green time ratio for a traffic light, to the geometric design of freeway interchanges.

These two classes of techniques, however, have been inadequate for the purposes of planners and traffic engineers with problems intermediate in scope between the region-wide network and the individual intersection. Typical of these "grey" areas are the study of detailed traffic movements within central business districts under peak-hour and off-peak conditions and within institutions such as airports and universities.

In such areas, region-wide assignments neither treat trip patterns or network geometry in sufficiently fine detail to give useful answers, nor do they attempt to respond to such factors as intersection control, parking regulations, or delays caused by congestion at intersections. On the other hand, the "behavioristic" models consider neither the impact of region-wide travel patterns on the study area nor the changes in vehicle routings through the study area caused by the buildup of congestion on certain links.

It was with a view to narrowing this gap in the technology of transportation planning that the Bureau of Public Roads (BPR) in 1967 retained the firm of Creighton, Hamburg, Inc., to develop a model designed specifically for traffic studies in small areas. Under the terms of the contract, the model was required to give an explicit treatment of all traffic movements in an area equivalent to approximately 200 city blocks and to provide data on link volumes, congestion delay, and travel costs for given time periods throughout the day. Such a "micro-assignment" model has been developed and is in operational use. It is implemented by a set of computer programs for the IBM System/360 which are now part of the Bureau of Public Roads Urban Transportation Program System.

The current version of the micro-assignment model permits simulation within a study area (micro-area) of up to 1,000 city blocks. Within the micro-area, every segment of highway network and each traffic movement through an intersection may be represented by a separate link. Furthermore, any network node (usually representing a block front) may be used as an origin or destination for trips.

For purposes of simulation, the user may divide the day into a number of time periods in any way he chooses. For each period, the model assigns to the micro-

Paper sponsored by Committee on Origin and Destination.

area network the proportion of the average total daily trips that occur during that period. During the assignment, the flow of traffic is allowed to increase gradually on the network. At regular intervals (usually after each minimum-path tree is built and loaded), the model computes the delay at each intersection caused by this buildup of traffic. The computed delays are then used to update the link travel time, which in turn influences subsequent minimum-path routings.

At the end of each time period, the program writes out a table of link volume and delays from which operating costs and network travel statistics can later be derived.

DESCRIBING THE MICRO-AREA

Size of the Micro-Area

The model places limitations neither on the shape of the area to be simulated nor on the configuration of the highway network within it. The present version of the model is designed to accommodate a network of up to 4,000 nodes and 12,000 links, equivalent to an area of approximately 1,000 city blocks. In the interest of economy in computer running time, however, and because of the computer memory requirements of the model, a more practical size for the micro-area seems to be from 200 to 300 blocks.

Network Layout

The main requirements of the micro-area network description are, first, that it accurately portray all the movements possible through each intersection (and only those movements), and second, that it contain all the data necessary to compute delays at each intersection.

To meet these needs it was found desirable to depart from the conventional method of representing networks as a collection of nodes (intersections) connected by links (road segments) to one in which each movement through an intersection is represented by a separate (one-way) link, with the road segment leading to the intersection treated as an integral part of the movement. Under this conception a link shares certain characteristics (such as speed, length, and parking restrictions) with the other links on the same approach leg, whereas other characteristics (such as number of lanes) are peculiar to the movement involved. A diagram of a typical four-way intersection, illustrating this concept of network layout, is shown in Figure 1.

Although this method of network layout requires more links than conventional representation, it overcomes two of the latter's major drawbacks. First, U-turns and other prohibited movements are automatically prevented from occurring during the minimum-path algorithm without recourse to "turn prohibitors"; and second, the need to compute turning volumes disappears.

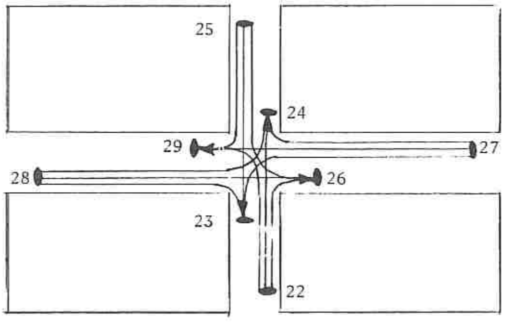


Figure 1. Conceptual layout of a typical intersection.

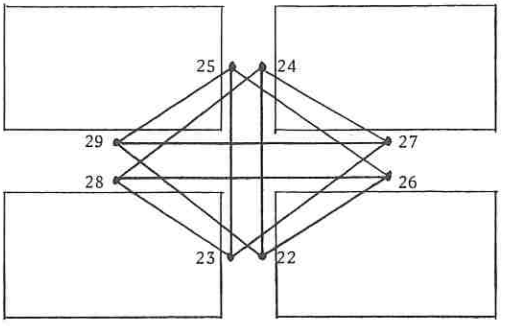


Figure 2. Intersection layout used for mapping.

Although the layout shown in Figure 1 is correct from a conceptual point of view, it has been found to be more convenient for purposes of mapping and node numbering to locate the network nodes at midblock, resulting in the intersection layout shown in Figure 2.

Because of the complexity of the intersection delay algorithm, it has been necessary to limit the number of movements leading out of a node to three—left, right, and through. An intersection at which more than three distinct movements are possible from any one approach leg represents a special coding problem. A method of dealing with this problem is explained in detail in the Creighton, Hamburg, Inc., final report (1).

DETERMINING TRAVEL IN THE MICRO-AREA

One of the principal inputs to the model is a file of trip interchanges that use the road network within the micro-area. Except in such applications of the model as uni-

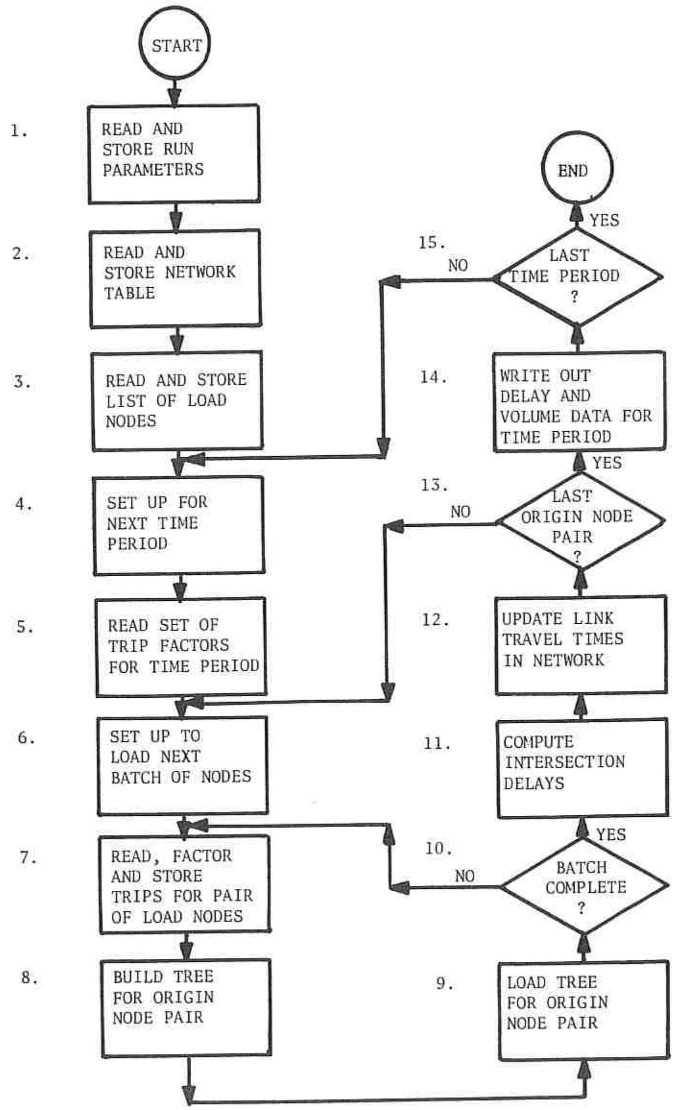


Figure 3. The micro-assignment process.

versities or airports where special travel surveys may be necessary, it is assumed that this trip file will be derived from existing region-wide origin-destination (O-D) tables.

This is done by use of a modified version of the BPR network loading program. One of the inputs to this program is a list giving every link in the region-wide network that crosses the micro-area boundary. The program then keeps a record of every O-D pair that uses the micro-area network for some portion of its minimun-path length, and notes the point(s), if any, at which these movements cross the boundary.

The output of this program may be used to select from a file of region-wide trip survey records all those that contribute to the micro-area trip population and to recode their origins and destinations to conform to the micro-area node numbering scheme. A further program step uses these selected survey records to create trip tables for use in the micro-assignment. This same program also prepares frequency distribution tables of travel purpose, which give the percentage of average daily traffic (ADT) trips entering the micro-area network during each 6-minute interval throughout the day. These tables are used to specify the percentage of trips of each purpose to be assigned during each time period selected for micro-assignment simulation.

THE MICRO-ASSIGNMENT PROCESS

The process described in this section involves a set of computer routines which, functioning as a single program, read the micro-area network description and trip tables and produce tables of link volumes and other network operating characteristics for selected time periods throughout the day. A flow diagram of this process (Fig. 3) is provided to supplement the text.

1. The first step in the micro-assignment process is to read in the parameters for the particular run. These are used to determine the amount of core storage needed for certain tables and to specify program user options. They include the highest node number, the number of network links, the number of time periods, and the number of travel purposes to be used.
2. The link data for the micro-area is read in and edited, and the internal network table is built. The program indicates any links containing errors.
3. A list of the nodes to be used as trip origins (load nodes) is read in and stored. The network loading routine causes minimum-path trees to be built and trips to be loaded only for those origin load nodes in this list, regardless of the constitution of the trip file being used. In this context, a load node implies an even/odd pair of network nodes, usually designated by the even node number (Fig. 2).
4. In preparation for the assignment of trips for the current time period, the link volume accumulation area is cleared and the travel time in each link of the network is set to its zero-volume value (i.e., to the value it would have for a solitary vehicle using the link). The program then is initialized to process the first node in the load node list.
5. Next, the program reads the set of trip factors (one for each purpose of travel) that determines the proportion of trips to be loaded during the current time period.
6. The user is given the option of building and loading a given number (batch) of trees before the program computes intersection delays and updates the link travel times. In this step the batch count is initialized to the value specified by the user.
7. The program now locates on the input tape the set of trip tables for the current load node and reads it into core storage. Tables for travel purposes not specified by the user are ignored. The trips for each node-to-node pair in the selected trip tables are multiplied by the appropriate factor and are accumulated into a single table for the current load node. If separate tables exist on the input tape for each of the nodes of the current even/odd pair, they are combined at this step.
8. A minimum-path tree is built for the current load node. Both nodes of the even/odd pair (if both exist) are regarded by the algorithm as though they were a single origin point.
9. In this step the trips from the current load node to all other pairs of nodes in the network are loaded onto the links of the minimum paths connecting them. Because

the completed tree always contains a different minimum path to each node of a destination pair (one approaching from each direction on the block) and because it is assumed that both sides of the street are equally accessible from either node of the pair, the trips are loaded onto the shorter of the two routes.

10. At this point the program tallies the number of trees loaded in the current batch. If the batch is not yet complete, the program returns to process the next load node (step 7). Otherwise it proceeds to step 11.

11. Using the duration time of the current time period, the program now converts the traffic volume on each link into the rate of arrival at the intersection (in vehicles per second) for that link. If the volume on a link was not changed during the loading of the current batch, an indicator is set for that link because it may not be necessary to recompute its delay.

The network is scanned to find those links with delays that must be recomputed. For each such link the program examines its characteristics (movement type, control device, etc.) to determine which form of delay computation is to be used.

12. As the intersection delay is computed for each link, the link travel time in the network is augmented by the amount of the delay.

13. When every link in the network has been examined and (if necessary) its delay computed, the program determines if the last origin node pair has been processed. If not, the program returns to step 6. Otherwise, the assignment for the current time period is complete, and the program proceeds to step 14.

14. The loaded network table (link volume) and the set of link delays in effect at the end of the time period are now written onto the output tape for later processing.

15. If the last time period of the run has been processed, the micro-assignment process is complete. Otherwise, the program returns to step 4 to set up for the next time period.

MICRO-ASSIGNMENT OUTPUT

The principal output of the micro-assignment model is a table for each selected time period giving for each link in the network a complete description of the link characteristics (used as input to the model) plus selected items computed during the assignment. These include computed delay, vehicle arrival rate and volume, operating speed and cost, and vehicle-miles and vehicle-minutes of travel. Additional tables for each time period summarize these data by type of intersection control, direction of movement (left, right, or through), and by number of lanes. Other tabular outputs from the model include listings of link data and tabulations of minimum-path trees and trip tables.

Graphic output from the model in the form of CalComp plots includes plots of trees, networks, and link volumes.

DELAYS AT INTERSECTIONS

A key feature of the micro-assignment model is a set of delay calculations detailed enough to be sensitive to such items as parking restrictions, number of lanes, signal timing, left-turn movements, and volume congestion and yet is simple enough to permit the simulation of traffic flow over an entire network. A collection of travel time and delay formulas were developed that met these conditions. A brief description of these equations and the methods of derivation are given in this section.

As a first step in analysis, travel time is decomposed into three parts: free travel time, zero-volume delay, and volume delay.

Free Travel Time

Free travel time is the time required to traverse a straight segment of roadway equal in length to the link length when traveling at the speed limit. Hence, this time is the link length divided by the speed limit. Travel time is computed once for each link and is not altered during the program.

Zero-Volume Delay

Zero-volume delay is the travel time in excess of free travel time which a single vehicle using an otherwise empty network would expect to encounter. This time includes delays caused by acceleration and deceleration for turns, stop signs, yield signs, and the random encounters of red lights and included the stop time for red lights.

Zero-volume delays are developed for each allowable intersection control device. These devices are

1. Yield sign;
2. Stop sign, cross facility not controlled;
3. Stop sign, cross facility controlled by stop signs;
4. Traffic signal;
5. Priority, local street (no control device); and
6. Expressway, limited access facility (no control device).

When the control device is a traffic signal, the zero-volume delay is further subdivided into type of movement: right turn, left turn, and through.

The only assumptions needed to derive these delays is that the vehicle uses uniform acceleration, deceleration, turn velocity, and cruising velocity. Introducing some notation, we let

a_1 = deceleration,
a_2 = acceleration,
v = turn velocity,
E = time needed to inspect intersection at a stop sign,
V = free flow velocity,
A_1 = time loss caused by stopping (through movement),
A_2 = time loss caused by acceleration and deceleration for turn (turn movements),
A_3 = $A_2 - A_1$,
R = red time,
G = green time,
C = cycle time, and
r = (G/C).

The use of elementary physics and mathematics yields the values of A_2 and A_3 and gives us the delay values given in Table 1.

$$A_2 = [(V - v)^2/(2V)]\,[(1/a_1) + (1/a_2)]$$

$$A_3 = [(v/2V)(2V - v)]\,[(1/a_1) + (1/a_1)]$$

TABLE 1

ZERO-VOLUME DELAY

Control	Delay Name	Formula
Signal (through)	δ_1	$r[(R/2) + A_1]$
Signal (right)	δ_2	$A_2 + r[(R/2) + A_3]$
Signal (left)	δ_2	$A_2 + r[(R/2) + A_3]$
Stop (2-way)	δ_5	$A_1 + E$
Priority	δ_6	0, through A_2, turn
Stop (4-way)	δ_7	$A_1 + E$
Yield	δ_8	A_2
Expressway	δ_{10}	0

Volume Delay

Volume delay is the travel time on a link attributable to the presence of other vehicles on the same link, cross links, and opposing links. This time includes

1. Time in queues;
2. Time waiting for suitable gap in the traffic stream;
3. Reaction time;
4. Conflict losses caused by traffic lane, intersection space, and green time being shared by vehicles making different but interacting intersection movements.

Volume delays have been developed for each control device. A brief description of each of these delay functions is given in the following sections.

Volume Delays at Signalized Intersections

Because of the complexity of vehicular movements at a signalized intersection, it was necessary to develop the following five categories of volume delay for this type of control:

1. Unimpeded through,
2. Unimpeded turn,
3. Left turn with interference from the opposing traffic stream,
4. Through with impedance from left turn vehicles in the through lane, and
5. Through with unimpeded turn movements in the same lane.

To derive expressions for these delays, we will need additional assumptions relative to the behavior of vehicles. The following characteristics are assumed:

1. Vehicle arrivals per cycle are Poisson distributed.
2. Vehicle arrivals during each cycle are uniformly distributed.
3. Vehicle size and queue spacing are uniform.
4. Reaction time is constant for all vehicles in a queue with the exception of the first vehicle.

These assumptions together with the one listed under zero-volume delay are sufficient to determine formal mathematical expressions for each volume delay element.

The uniform arrival of vehicles during any given cycle is used to derive delay expressions as a function of demand (arrivals/cycle). The Poisson distribution of arrival rates then is used to compute the expected volume delay by multiplying the delays derived for a particular demand by the probability of that demand after "s" cycles and summing over all possible demands. The value s is the final cycle after a demand persists for 10 minutes. It is computed by dividing the cycle length into 600 seconds and rounding to the nearest integer.

The bookkeeping involved in recording the changing queues and demand possibilities through several cycles makes this a cumbersome process. Hence, it is used only as an intermediate step to reach a form capable of being computed more rapidly. For example, the process leads to the following delay expressions for unimpeded through traffic:

$$L_1 = \sum_i L_1(i)\, P_S(i)$$

where

$i = 0, 1, 2, \ldots;$

$$P_S(n) = \sum_{r=0}^{n} P_{S-1}(n + K_1)\, P_{S-1}(n - r)$$

and where

$P_1(n) = (1/n!)\,(VC)^n \exp(-VC);$

K_1 = capacity/cycle;

V = demand volume (Poisson distributed); and

C = cycle length.

The expressions for $L_1(i)$ are given in Table 2 where

K_1 = capacity of through movement,

$d(n)$ = average time lost by a vehicle in a queue of n vehicles because of reaction times,

$L_1(n) = (R/C)\,[d(nR/C)] + [(nRG)/(C^2K_1)]\,\{[(nRG)/(2CK_1)] + A_1 + d[(n^2RG)/(C^2K_1)]\}$ for $n \le K_1$.

The delay function L_1 when expanded back to original input variables becomes very cumbersome. This difficulty is avoided in practice by using a cycle-to-cycle iterative process.

TABLE 2

DELAYS FOR UNIMPEDED THROUGH TRAFFIC

Demand (number of vehicles)	Expected Delay
1	L_1 (1)
2	L_1 (2)
⋮	⋮
K_1	L_1 (K_1)
$K_1 + 1$	$L_1 (K_1) + [1/(K_1 + 1)] [R + A_1 + d (1) - L_1 (K_1)]$
$K_1 + 2$	$L_1 (K_1) + [2/(K_1 + 2)] [R + A_1 + d (2) - L_1 (K_1)]$
⋮	⋮
$2K_1$	$L_1 (K_1) + [K_1/(K_1 + K_1)] [R + A_1 + d (K_1) - L\ (K_1)]$
$2K_1 + 1$	$L_1 (2K_1) + [1/(2K_1 + 1)] [C + R + A_1 + d (1) - L_1 (2K_1)]$
$2K_1 + 2$	$L_1 (2K_1) + [2/(2K_1 + 2)] [C + R + A_1 + d (2) - L_1 (2K_1)]$
⋮	⋮
$3K_1$	$L_1 (2K_1) + [K_1/(2K_1 + K_1)] [C + R + A_1 + d (K_1) - L_1 (2K_1)]$
$3K_1 + 1$	$L_1 (3K_1) + [1/(3K_1 + 1)] [2C + R + A_1 + d (1) - L_1 (3K_1)]$
$3K_1 + 2$	$L_1 (3K_1) + [2/(3K_1 + 2)] [2C + R + A_1 + d (2) - L_1 (3K_1)]$
⋮	⋮
$4K_1$	$L_1 (3K_1) + [K_1/(3K_1 + K_1)] [2C + R + A_1 + d (K_1) - L_1 (3K_1)]$
$4K_1 + 1$	$L_1 (4K_1) + [1/(4K_1 + 1)] [3C + R + A_1 + d (1) - L_1 (4K_1)]$
⋮	⋮

Although the computed expected delay yields satisfactory results, its use in an assignment program where hundreds or even thousands of such calculations must be made would result in long, expensive programs. To avoid this contingency an empirical equation was developed that approximates the previous results and is capable of rapid evaluation.

Volume Delays at a Stop Sign, Cross Facility Not Controlled

This delay is composed of two parts: Time waiting at the stop sign for a suitable gap in the cross traffic, and time waiting in a queue to reach the intersection when a queue exists.

To account for the delay of waiting for a suitable gap in the cross stream of traffic, heavy reliance was placed on the work of others. Gazis et al. (2) have developed expressions for the expected waiting time at a stop sign by a vehicle that is trying to cross a road with n lanes, given the vehicles' gap acceptance characteristics (acceptable gaps for each of the n lanes) and the mean arrival rates of the n lanes. An approximation formula is also developed that uses the sum of the mean arrival rates and a single weighted gap acceptance value. Because this formulation is more in keeping with data bases for an entire area, it has been retained for use in the microassignment program. Let

$F(i)$ = mean arrival rate on the ith lane;
$T(i)$ = gap acceptance time for the ith lane; and
b = expected delay time.

If

$$F = \sum_i F(i); \text{ and}$$

$$T = \left[\sum_i F(i)\,T(i)\right] \Big/ \sum_i F(i);$$

then the required expression is

$$b = (1/F)\,[\exp(FT) - 1] - T$$

This same expression has been developed under different assumptions by other researchers (3, 4).

To find the time lost in queues, the service time γ is defined as the reciprocal of gap time b; the service time thus is exponential in form. Because the vehicle arrivals are assumed to be Poisson distributed, we may turn to queuing models for a single lane to find an expected delay of $V/(\gamma^2 - \gamma V)$.

Because this service is based on arrival rates persisting indefinitely, it gives unbounded delays as V approaches γ in value. We are interested, however, in shorter time intervals, and therefore it seems logical to bound this expression by a delay that would result from the demand persisting for 10 minutes: $\Phi + 600\,(V - \gamma)/(2\gamma)$, where Φ is chosen to make a smooth tangent with $V/(\gamma^2 - V)$. The point of tangency v can be found by equating slopes:

$$\frac{300}{\gamma} = [1/(\gamma^2 - \gamma v)] + [\gamma v/(\gamma^2 - \gamma v)]$$

making

$$v = \gamma - (\sqrt{3\gamma}/30)$$

This implies that

$$\Phi = [(20\sqrt{3\gamma})/\sqrt{\gamma}] - (1/\gamma)$$

Hence, the expected delay L_5 is equal to

$$b + [V/(\gamma^2 - \gamma V)] \text{ for } V \leq v$$

and

$$b + \Phi + \{[300\,(V - \gamma)]/\gamma\} \text{ for } V > v$$

Volume Delays on Local Streets, No Intersection Control

The computation of these delays can get somewhat involved because of the various movements sharing lanes and because of frictions arising from high densities on the lanes. The latter effect is difficult to deal with and is not dominant in systems where priority links are coupled with links having some other intersection control. Because this is true in most systems of interest, the volume delays discussed are limited to those arising from lane sharing.

Through vehicles with their own lane experience no delay whereas those that share a lane with a turning movement do experience some delay. To simplify the delay calculation for those movements, through vehicles are assumed to suffer delays in accordance with the least hampered of the through lanes.

For the through movement with its fastest lane shared with a right turn movement or a left turn movement without interference, this delay is just A_2.

For a left turn movement with interference (the usual case because the opposing traffic commonly is also a priority movement), the delay is the same as for the burdened leg except that a full stop is required only as dictated by opposing traffic. Ignoring this small difference, we may use L_5 to represent this delay.

The values of L_6 are summarized as

A_2, for through movement when this fastest lane is shared with right turn or left turn with no interference;

L_5, for left volume; i.e., for left turn with interference or for other movement sharing its lane with left turn with interference; and

0, for through movement (at least one lane unshared), for right turn with no interference, or for left turn with no interference.

Volume Delays at a Stop Sign, Cross Street Also Controlled by a Stop Sign

The calculation of this volume delay could get quite involved. Because there are typically 12 movements at a four-legged intersection, random arrivals give rise to many vehicular arrangements and the various possibilities cannot be ordered as neatly as in the case of signal-controlled intersections. To find the probability of all possible arrangements, the 12 mean arrival rates and the associated average delay for the 12 movements must be given and then the sum of the products of probability times delay for all movements must be found for all possibilities. This task is not feasible for network analysis.

To reduce the problem, assume for the moment that all the vehicles on the various legs form a single serving line with a mean service time (capacity K_7) equal to the reciprocal of usage time (which we assume to be exponentially distributed). We may then apply the Poisson single-channel-arrivals queuing model to find the expected delay per vehicle at the intersection. By bounding this expression to the delay that would result after the demand persisted for 10 minutes and by prorating the resulting delay among the various links using the intersection, we find the following expected delay on the i th approach leg:

$$D(i) = [V(i) \Sigma V(i) L_7]/\Sigma [V(i)]^2$$

The summation is taken over all of the approach legs to the intersection and

$$L_7 = \Sigma V(i)/[K_7^2 - K_7 \Sigma V(i)] \quad \text{for } \Sigma V(i) \leq v$$

and

$$L_7 = \Phi + [300 (\Sigma V(i) - K_7)/K_7] \quad \text{for } \Sigma V(i) > v$$

where

$$\Phi = (20\sqrt{3}/\sqrt{K_7}) - (1/K_7)$$

Volume Delay at a Yield Sign

The zero-volume delay for links with this control device is somewhat less than that of the stop-sign facilities (Table 1) because the vehicle need only slow down when no other vehicles are present. The delays caused by other vehicles are those of waiting in queues and waiting for a suitable gap in the cross traffic. Because these are the same elements that occur on facilities controlled by stop signs, the volume delay for a link controlled by a yield sign is the same as that of a stop-sign-controlled link.

Volume Delays on Expressways

The volume delay developed for expressways is quite simple and does not involve the problem of unstable flow. There may be questions, however, about the method of handling demand in excess of capacity. On the other hand, the treatment of delays for demands not exceeding the capacity are in line with the studies and observations of previous researchers. Also, for the case of demand in excess of capacity, the volume delay increases rapidly; this is certainly the correct behavior even if there is some question about the absolute magnitude of these large delays.

The delays are derived as follows. Speeds are reduced linearly from the speed limit at zero demand to one-half the speed limit at a demand equal to the capacity. For demands exceeding the capacity, the speed continues to decrease linearly with the amount the demand exceeds the capacity. This implies that queues form on the entrance ramps. The existance of these queues increases drastically the average delay. Also implied by this assumption is a flow rate of zero when the demand is twice the capacity. This of course implies an infinite average delay per vehicle. Although there actually might be some small nonzero flow at this demand, it would not be much greater than zero and the authors are not unduly alarmed at the asymptotical behavior of the

derived delay function. The translation of this into mathematical equations is not particularly difficult. The results are

$$L_{10} = [\ell V_{10}]/[v\,(2K_{10} - V_{10})] \text{ for } V_{10} \leq 2K_{10}$$

$$L_{10} = \infty \text{ for } V > 2K_{10}$$

where

v = speed limit,

ℓ = length of expressway link,

K_{10} = capacity of a lane of expressway, and

V_{10} = demand volume on the expressway link.

Applications of the Volume Delay Formulas to Specific Intersection Configurations

To correctly apply the delay functions, the number of lanes available to a link must be known. This determination is complicated by the presence of parked vehicles, turn bays, and various intersection movements sharing one or more lanes. From the link description on one approach leg and the internal clock (parking restrictions, turn prohibitions, and number of lanes vary by time of day), a subroutine called "intersection switching" determines the number of lanes available for each movement and the number of these lanes that must be shared with other movements. This information, together with certain assumptions as to vehicular arrangement on shared lanes, enables the computer to select a final expected delay for the link that is a weighted average of the expected delays discussed previously. Space does not permit a development of these final expected delays in this paper, but Figure 4 shows a diagram of each approach-leg configuration that can be handled by the program at this time.

TESTING THE MODEL

When the programs written to implement the model were considered operational, two sets of tests were made to evaluate the performance of the model. The first of these was designed to assess the behavior of the delay computations for various intersection configurations and under various traffic loadings. The second set of tests involved making assignments over an actual downtown area (Buffalo, New York) so that certain results of the model might be compared with observed data.

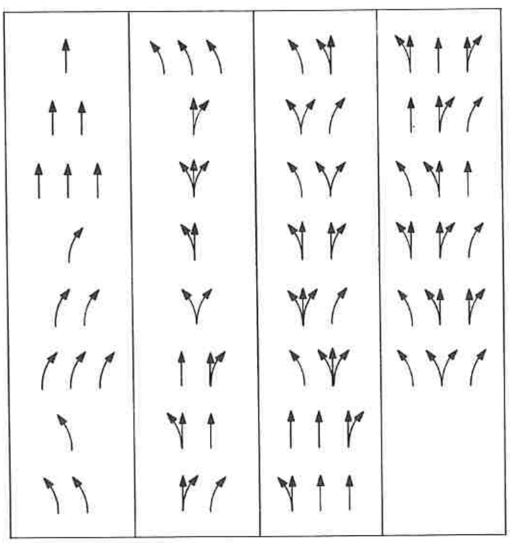


Figure 4. Possible approach-leg configurations for signalized intersections.

The first set of tests was run using a hypothetical network consisting of a string of 19 intersections; 15 of which were controlled by traffic lights; two, by four-way stop signs; and the remaining two were through streets, with stop signs controlling the intersecting streets.

The number of lanes, the total cycle time, the green cycle time, and the arrival rates were varied in a systematic manner for each of five time periods, during which various parking and movement restrictions were imposed.

A complete tabulation of the results of these test is included in the study's final report (1), and a few of them are discussed here for illustrative purposes.

Table 3 gives the arrival rates and parking restrictions in effect during four of the test periods; Table 4 gives the delays computed for the through movements at each of six intersections for each time period. Delay times and cycle times are both in seconds. It should be noted that these delays are in addition to the zero-volume delays, which are generally on the order of one-half the duration of red time for each link.

TABLE 3

CONDITIONS PREVAILING DURING THE TEST TIME PERIODS

Time Period	Parking	Arrival Rates (vehicles/min)		
		Through	Left	Right
1	Yes	12.0	1.2	1.2
2	No	12.0	1.2	1.2
3	No	24.0	2.4	2.4
4	No	0.6	0.3	0.3

The effect of allowing parking can be seen by comparing the delays for period 1 with those for period 2 (Table 4). In particular, links 1 and 3 exhibit marked increases in delay when lanes were effectively reduced to one. Link 5, with its greater green time ratio, is much less affected by the loss of a lane; and movements that shift from three to two lanes because of parking do not experience much increase in delay in the process.

The arrival rates used for period 4 were apparently too small to cause any significant delay (over and above zero-volume delay).

A comparison of the delays for periods 2 and 3 shows the effect of doubling the arrival rates on each link. Although links 2, 4, 5, and 6 still remain well below capacity in period 3, the volume on link 1 has exceeded the discharge capacity of the intersection, resulting in an average delay of nearly two full signal cycles. Link 3, under these conditions, shows an average delay of only half a signal cycle, reflecting the greater capacity obtained by using this signal setting.

TABLE 4

DELAYS ON THROUGH LINKS FOR EACH TIME PERIOD

Link	Lanes	Total Cycle	Green Cycle	Congestion Delay Time (sec) for Time Period Number			
				1	2	3	4
1	2	60	30	150	3	104	0
2	3	60	30	3	2	4	0
3	2	120	60	118	6	60	0
4	3	120	60	6	4	8	0
5	2	120	90	7	3	5	0
6	3	120	90	3	1	4	0

TABLE 5

SUMMARY OF RESULTS FOR BUFFALO, NEW YORK, MICRO-AREA

Time Period	Trip Length		Average Speed (mph)	Percent Vehicles Turning		
	Vehicle-Hours	Vehicle-Miles		Left	Through	Right
5:00-6:30 a.m.	329	3,304	14.4	12.9	70.4	16.7
6:30-7:30 a.m.	564	8,608	15.2	11.9	72.4	15.7
7:30-9:00 a.m.	2,409	26,891	11.1	12.2	72.0	15.8
9:00-10:00 a.m.	712	10,272	14.4	11.9	71.9	16.2

For the second series of tests, a micro-area network was prepared for a portion of downtown Buffalo, New York. In the absence of much of the detailed data needed to complete the network, many link and intersection characteristics had to be estimated. The trip files in these tests were derived from the 1962 O-D survey conducted by the Niagara Frontier Transportation Study (NFTS).

The main purpose of this series of tests was to study the behavior of the model in a real downtown area. Although no traffic flow data were available by time-of-day with which the model's results might be compared, it was hoped to find agreement on some of the broader parameters of network performance.

The results obtained from four time periods, which included the morning peak period, are given in Table 5.

The average speeds obtained in this run are substantially in agreement with those obtained in the NFTS driving time survey (16 mph during off-peak hours), and exhibit the decrease expected during the height of the morning peak period. The percentages of turning vehicles also agree very well with those obtained in other studies.

Although the number of tests made so far on the micro-assignment model is insufficient to be the basis for any dramatic claim on its behalf, the results must be regarded as an encouraging indication of its potential.

REFERENCES

1. Creighton, Hamburg, Inc. Micro Assignment—Final Report. Bureau of Public Roads Contract FH 11-6755, Washington, D.C., 1969.
2. Gazis, D. C., et al. The Delay Problem for Crossing an n-Lane Highway. Proc. Third Internat. Symp. on the Theory of Traffic Flow, 1965. Elsevier, New York, 1967.
3. Greenshields, Bruce D., and Weida, F. M. Statistics—With Applications to Highway Traffic Analysis. Eno Foundation for Traffic Control, Saugatuck, Conn., 1952.
4. Adams, W. F. Road Traffic Considered as a Random Series. Jour. Inst. Civil Engineers (London), Vol. 4, 1936.

Introducing the Idea of the "K Distribution" to Transportation Patterns

YACOV ZAHAVI, Kolin and Zahavi, Transportation and Traffic Engineers, Tel-Aviv, Israel

The "Distribution of Factor K" is an empirical distribution that has been found to operate in many research fields. The application of this distribution by the transportation engineer may contribute considerably to the definition of some important aspects of transportation. This empirical law should help solve, for instance, the problem of defining an index of the pattern of concentration of traffic, and of the relationship between the traffic-attraction powers of the various traffic zones in a city. Another problem is that of defining the scope of the significant data essential to planning. The transportation planner often is confronted with an abundance of data relating to numerous zones. A great deal of time and expense could be saved were it possible to confine the scope and limit to the most significant data. It is the object of this paper, then, to present a possible approach to these and other questions by applying an extended form of the "K Distribution" to some transport behaviors.

• IN VARIOUS RESEARCHES and in different spheres, the following empirical phenomenon has been noted. When ranking certain observations in a decreasing order of their value, where the rank of the observation is plotted along the x-axis and its value along the y-axis, a straight line is formed on a double logarithmic scale. This straight line can be formulated as follows:

$$r = KP^{-q}$$

where

r = the rank of the observations,
P = the value of the observations,
q = an exponent, the value of which nears 1 as the slope increases, and
K = a constant factor, the value of which is close to the highest observation.

The formation of such a straight line means that the values of the observations decrease in a regularity that is bound to a certain relationship between the larger observations and the smaller ones. Values decrease sharply at the beginning; later, the decrease becomes smaller and smaller. Such a hyperbola-like distribution provides the points on the top of the straight line with much greater weight than the ones at the bottom of the line.

The slope of the line expresses the degree of the relationship between the observations, or the level of differentiation in their weights of activity. A moderate slope, while still possessing its hyperbola-like distribution, denotes a smaller such differentiation.

Paper sponsored by Department of Urban Transportation Planning.

On the log-log scale, some of the observations are situated at the margin of the straight line; their values are low and they do not continue the trend of the straight line. Thus, a breaking point of the straight line is located that terminates the above described specific regularity of the observations.

The K distribution (KD) thus briefly described has been observed in a number of demographic, sociologic, and biologic behaviors. Certain alternative explanations have been given to it, like those offered by Zipf, Rashevsky, Simon, and others. This paper is not going to offer another one but, as stated, is going to apply its empirical operation for practical uses in some transportation issues.

Before this application is discussed, it would be useful to note the operation of the KD in some other fields and to evaluate its meaning concerning those specific variables.

The KD has been repeatedly noted in the size distribution of settlements. Table 1 and Figure 1 show these distributions for the United States and Israel (1). In both cases a straight line is formed and broken at the level of the smaller settlements. The KD is applicable to the settlements for different countries and thus has served as a basis for some significant demographic conclusions.

1. The straight line indicates that the settlements situated along it are developing in a specific regularity, different from the regularity governing the rest of the settlements. Their varying size is related to a specific order so that their power to attract new inhabitants is in proportion to their relative size. Figure 1 shows that this regularity is common to both very large and very small countries.
2. The slope of the line defines the "level of urbanization" of a country. When the slope is steep, like that for the United States and Israel, a major part of the population is concentrated in urban settlements. A moderate slope, on the other hand, signifies a dispersion of the population and expresses a less extreme differentiation between the settlements and thus a low level of urbanization. These moderate slopes are typical of agricultural or nomadic countries. Another characteristic of the slope is that the steeper it gets, the fewer the settlements along the straight line and the more below the breaking point.
3. Demographers have defined settlements along the straight line as "urban settlements." Thus, the breaking point of the line separates the larger settlements ranked as urban settlements of the country, from the settlements which, at that time, exert a weaker power of attraction and constitute a smaller proportion of the population. The point at which a settlement becomes urban lies on different levels in different countries. In the United States or Israel, for instance, the breaking point is at a high level; a settlement requires some 10,000 inhabitants (Fig. 1) to be urban. Within the hierarchy, however, there are countries in which the settlements become urban at a much lower

TABLE 1

POPULATION IN LOCALITIES BY SIZE-CLASS

Size of Locality (number of inhabitants)	United States of America 1.4.1960		Israel 22.5.1961	
	Number of Localities	Population	Number of Localities	Population
Total		179,323,175		2,179,491
In localities	19,790	125,808,073	873	2,148,310
500,000 and over	21	28,595,050	—	—
100,000-499,999	111	22,418,307	3	736,526
50,000-99,999	201	13,835,902	2	144,841
20,000-49,999	632	19,400,682	15	471,048
10,000-19,999	934	13,118,216	14	175,311
5,000-9,999	1,394	9,779,714	20	145,026
2,000-4,999	3,048	9,577,903	50	164,797
1,000-1,999	3,575	5,049,869	51	75,983
500-999	3,267	2,341,061	121	75,932
200-499	4,153	1,389,190	411	137,862
Under 200	2,454	302,179	186	20,984
In others		53,515,102		31,181

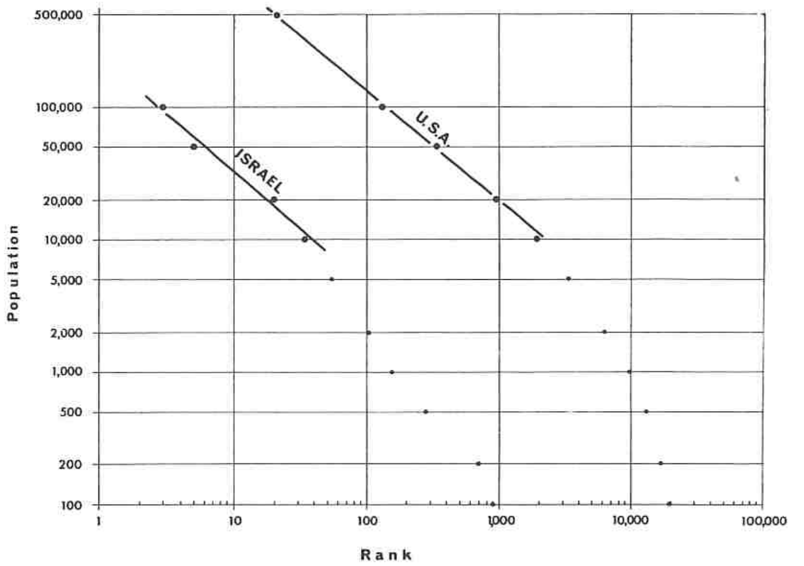


Figure 1. Distribution of settlements in the United States of American and Israel.

level. The breaking point marks the level above which settlements can be considered as dominant ones, while those beneath it can be considered as minor within the distribution.

Conclusions of this kind, which have been reached upon applying the KD to demographic studies, have been verified in other fields as well. To give another example, the KD can be found even in some quite unexpected spheres, as in the frequency distribution of the alphabetic letters in the Bible. In the 10th century, Rabbi Saadya Gaon calculated the frequency at which the different letters appear in the Old Testament, as given in Table 2. Plotting these frequencies by the method just discussed forms once again a straight line with the typical breaking point, as shown by Figure 2. Similar conclusions can be drawn from the distribution of the letters of the alphabet regarding the concentration of the use of the various letters, the regularity of their appearance, the dominant nature of certain letters, etc.

It is possible, therefore, that some important findings might emerge concerning a variable known to be K distributed. Consequently, the KD will be of special interest to transportation engineers, if it turns out that it operates for certain transportation distributions as well. Our findings show that this happens to be true.

Investigation of a series of data concerning trip-ends in different towns of various sizes showed that their distributions plotted on a log-log scale assume the typical form of the KD with surprising accuracy. From a large number of similar cases, three examples of such distributions are shown by Figure 3.

Figure 3 shows the manner in which the attracted trip-ends are distributed in Chicago (2), Pittsburgh (3), and Tel-Aviv (4). Typical to a KD, the main traffic zones or districts, which attract most of the trips, are situated around a straight line in all three towns. In all these cases, three

TABLE 2

THE FREQUENCY OF APPEARANCE OF ALPHABET LETTERS IN THE BIBLE

Letter	Frequency	Rank	Letter	Frequency	Rank
Alef	42,377	7	Mem	77,778	1
Beth	38,218	10	Nun	41,696	8
Gimel	29,537	13	Sameh	13,580	21
Daleth	32,530	11	Ayin	20,175	20
He'h	47,754	6	Pe'h	22,725	17
Vav	76,922	2	Zadik	21,822	19
Zayin	22,867	16	Kof	22,972	15
Heth	23,447	14	Resh	22,197	18
Teth	11,052	22	Sheen	32,148	12
Yod	66,420	3	Taf	59,343	4
Kaph	48,253	5	Total	815,330	22
Lamed	41,517	9			

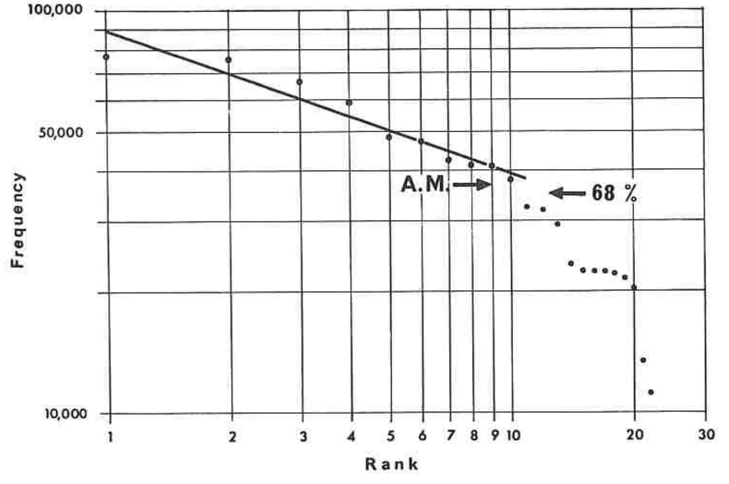


Figure 2. Frequency distribution of alphabet letters in the Bible.

different slopes can be observed. At relatively low levels of trip attractions a breaking point occurs in all three towns, though the number of trip-ends it signifies is different, of course, for each.

Examining the distribution of traffic volumes on the road network by the hours of the day showed that they too are K distributed. In Figure 4, public transport passenger trips in Tel-Aviv (4) are plotted on a log-log scale. Once again, characteristic to a KD, the hourly trips form a straight line with a specific slope, which is clearly broken at a low level of passenger trips.

These are only two examples of traffic distributions that can be considered to be K distributed. The deductions that the traffic engineer may arrive at by applying the characteristics of the KD to the attracted trip-ends distributions and to the distribution of traffic volumes by the hour cannot be overestimated. Before these deductions are dealt with, however, one characteristic of the KD, i.e., the breaking point, merits some special attention.

THE DOMINANT OBSERVATIONS

Whereas the empirical KD has been recognized and dealt with in numerous publications, no definition or location of the breaking point has been offered, and its significance has not yet been fully appreciated.

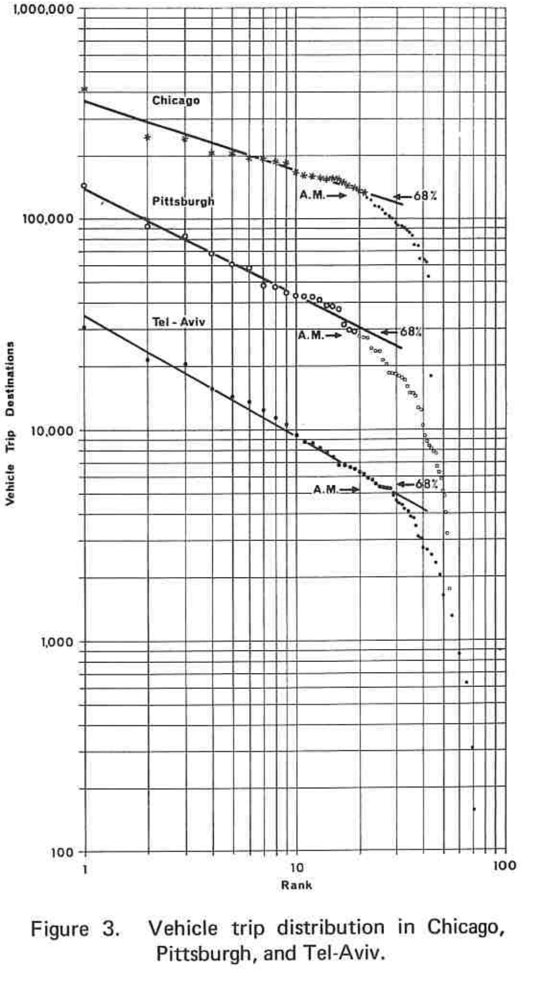


Figure 3. Vehicle trip distribution in Chicago, Pittsburgh, and Tel-Aviv.

In the framework of the Transportation Master Plan for Tel-Aviv (4), in which the operation of the KD in certain traffic distributions was found, the essential nature of the breaking point was thoroughly investigated. Although different sorts of trip-end distributions of various towns were examined, an interesting possibility emerged for a clear definition of the characteristics of the breaking point. This has been found to be true for all the K-distributed variables examined. Comparisons made to this end marked two different parameters by which the breaking point may be defined: (1) by the arithmetic mean of the observations; and (2) by the sum of observations, the value of which is close to 68 percent of the total.

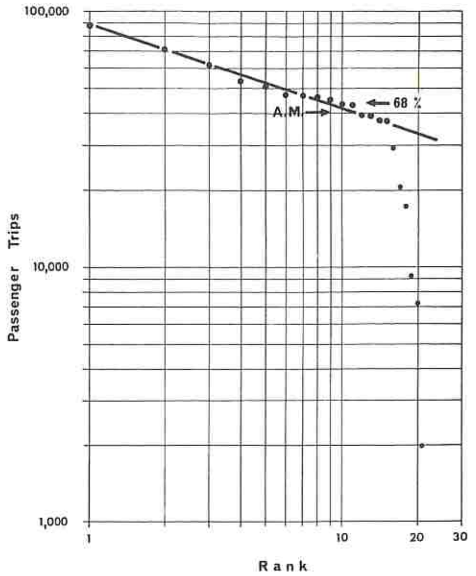


Figure 4. Distribution of bus passenger trips in Tel-Aviv, 1965.

1. It has been found that summing up the values of all observations and dividing their total by the number of observations, an average value is obtained which, surprisingly enough, marks the breaking point of the straight line. The average size of the settlements, for instance, has been found to indicate quite closely the point at which the settlements cease to follow the initial trend of the straight line. The traffic zone that attracts the average number of trips, as another example, marks in most cases the point at which the straight-line trend of the traffic zones that attract most of the trips is broken. This average stratifies the zones, then, into the dominant ones and others. Figures 2 through 4 show this relationship between the breaking point and the arithmetic mean (AM) of the observations. This established an empirical correlation that appears to be incidental but may well be reasonable and logical. One of the explanations describes the observations as being distributed around an average condition in a steady-state environment. The observations above this average are, therefore, more active than those below it. Moreover, the empirical distribution shows that the interrelationship characterizing the dominant zones, i.e., those above the average, is entirely different from that characterizing the observations below the average.

2. In most of the KD it has been noticed repeatedly that the value of the observations above the breaking point amounts to about 68 percent of the total value of the observations. This percentage reminds us of the observations that deviate from the mean, in a normal distribution, by one standard deviation. This might indicate that those observations that are dominant in character range by only an average deviation from the highest observation. This recurrent relationship seems, too, to be casual; essentially, however, the relationship demonstrates that the dominant and the active observations (the active towns or traffic zones) are those close in character to the highest and most active observation (the main town or traffic zone). Figuratively, by borrowing the statistical normal curve concepts, we can say that in a KD only those units having a total value that does not exceed a normal deviation are along the straight line and above the critical point.

It should be remembered that the KD itself is empirical and does not operate in all cases with mathematical precision. Therefore, the above definition of the breaking point should not be expected to pinpoint its exact location, but serves to define it with sufficient accuracy for actual planning purposes. In most of the K distributions examined, this relationship has been found to exist with surprising accuracy. Figures 2 through 4 show this relationship.

The two empirical factors somewhat extend the accepted notion of the KD and provide a better position from which to elaborate the significance of applying the KD to some transportation patterns.

THE K DISTRIBUTION—A DEVICE FOR TRANSPORTATION PLANNING

This section details some of the significances arising from applying the KD to some transportation distributions. The extended concept of the KD developed in the previous chapter has some new bearings on other fields, too, especially the demographic field. This section, however, will deal only with its significance concerning transportation issues.

One of the most important transportation distributions is that of the spatial dispersion of the produced and attracted trip-ends between the various traffic zones in a city or a metropolitan area. The significances of this application are discussed as follows, especially in connection with these distributions, although the significances are general in character.

Levels of Traffic Concentration

A major parameter characterizing the travel pattern of a city is its level of traffic concentration. If the distribution of trip-ends is said to be K distributed, then the slope of its straight line may serve as an indication of the traffic concentration in that city. A steep slope indicates that a great number of trips are attracted to some selected active zones whereas the rest of the zones attract considerably fewer trips. The steep slope is an expression of a concentrated pattern of trips or of a high level of "traffic urbanization" in that city. A moderate slope, or a low level of traffic urbanization, indicates that each of the many zones attracts relatively few trips and that there are no outstanding strong zones of destination.

Obviously, planning for such different conditions of travel will take quite different directions. The slope of the line, therefore, may assist in determining the specific planning approach to a specific travel pattern.

Development Trends

The slope of the straight line may be compared between periods of time. Such comparisons should point out the direction of the development trends of traffic urbanization of a metropolitan area. Before conducting such comparisons, however, we should review the essential nature and characteristics of the unique variations in the slopes of K distributions.

In early stages of a country's development, as is the case in agricultural or nomadic countries, the slope is quite moderate. In the next stage of development, with an outburst of growth of some settlements to the level of big towns, the slope becomes increasingly steeper and the level of urbanization increases accordingly. Surprisingly enough, in the next phase of development, the slope of the line may get moderate again. The reason is immediately apparent: At that stage the main growth continues in the rest of the satellite metropolitan towns because the central city has already approached a certain saturation level with high population density. The level of concentration, therefore, is decreasing; thus the slope returns to a more moderate form. An example of such a development is the growth that is taking place in the New York-Baltimore belt.

Parallel variations in the slopes are observed in the distribution of trips to zones of destination in a metropolitan area. At the beginning, the slope is moderate because the distribution of the trips is rather dispersed. The level of trip concentration increases as the main town of a metropolitan area develops and creates some major zones of destination. At that stage the metropolitan area is said to have a high level of traffic urbanization. At the next stage of variation in the travel pattern, a process of trip dispersion sets in again: The other satellite metropolitan towns start to attract an increasingly larger number of trips; but the main city, after reaching a certain level of saturation, does not develop at such a pace.

Bearing in mind the nature of the variations of the slope in a KD, we can gain considerable knowledge concerning the trends of traffic urbanization in a certain area by comparing its slopes for different periods of time.

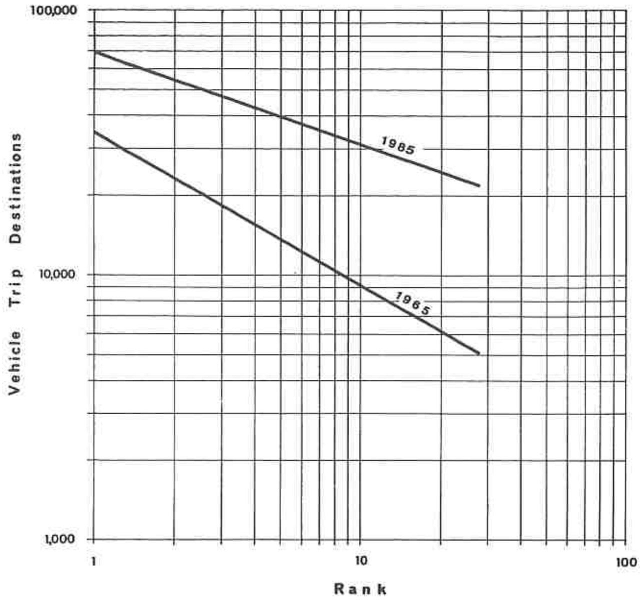


Figure 5. Vehicle trip distribution in Tel-Aviv, 1965 and 1985.

Figure 5 shows the trend of concentration in the distribution of trip-ends in the Tel-Aviv metropolitan area. The figure shows the distribution for 1965 and the one forecast for 1985. The change in the slopes is clearly noticeable. An examination of the meaning of the slope's movements leads to a clear conclusion concerning that change. The level of traffic urbanization, or traffic concentration, in the Tel-Aviv area is apt to decrease. The zones at the outskirts of the main city are going to develop considerably in a process of differential growth so that the dispersion of population, land uses, and trips will increase. A marked change in the weight of trip attraction will occur between the central and the satellite towns. This trend should have some significant bearings on future plans for the road network in the area.

It should be stressed at this point that for the purpose of comparisons between different towns or time periods, an equivalent definition of traffic zones should be maintained in each town or period. Different criteria for defining the zones may produce different slopes for the same area. Comparisons such as those shown in Figure 5 produce the best results because they refer to exactly the same zones in both periods.

Comparisons of Distributions

The slope of the straight line may be compared between different distributions as well. The comparison, for instance, between the attracted trip-ends for private vehicles and for public transport might produce some interesting results. The distributions in Tel-Aviv for 1965 and 1985 respectively were compared. The results were as follows:

1. The slope of the line for the public transport distribution was considerably steeper; the exponent of the slope had a value of 0.67, whereas the exponent for the private vehicles for the same year was 0.57.
2. With the general decrease in concentration by 1985, both distributions became less concentrated, but to a lesser extent for the public transport. The exponents for 1985 were 0.46 for public transport and 0.35 for private vehicles.

One of the most important conclusions to be drawn from these results is that a quite significant difference exists between the two distributions. The concentration for transit

trips is much more notable than for private vehicles; this trend increases relatively over the years, despite the general trend toward a greater dispersion of trips.

Upon applying models of trip distribution, by the way, it would be most desirable and useful to employ two separate models, one for each mode. This is warranted by the different pattern of centralization of the two modes. The method of employing different models for the different modes seems vastly preferable to the method that applies a single model for the total person-trips, at the risk of having to split it into two spatial ones, i.e., one for the central business district and one for the rest of the metropolitan area. The need for different models is more acute for towns with a high percentage of public trips. Separate examination of the trip distributions for the two different modes, in addition to an investigation of their trends by means of the KD, therefore, should provide a better understanding of the interrelationship between the two modes and should introduce a more reliable method of forecasting the modal-split ratios.

Variations in Traffic Zones

Examination of the traffic zones along the straight line may indicate, in addition to the general trend of trip concentration in a city, the variations in individual traffic zones as well. Traffic zones situated along the straight line may drop below the breaking point, and vice versa. This will be true with some traffic zones in Tel-Aviv by 1985. Such developments should have been expected because, as has been pointed out already, changes occur with time in the relative weight of zones. A change in the location of certain zones along the straight line indicates such a change in their power of trip attraction. A knowledge of the regularity with which progression or regression of the zones located along the straight line occur together with the knowledge of the direction and strength of the changes in the location of some of them may contribute considerably to the understanding of the process of relative change affecting the zones with the lapse of time.

Locations of Dominant Traffic Zones

One of the most useful consequences of applying the KD is that it facilitates the location of the breaking point in trip-ends and other transportation distributions. As has been seen, the observations plotted along the straight line were those of above the average value, those constituting a greater proportion (about 68 percent) of the total trip-ends. This makes possible an immediate and simple definition of the "dominant traffic zone." This definition can serve as a sufficient basis for a great many transportation plans; other zones may be disregarded because their data are not essential for the planning process.

The location of the breaking point in the trip-ends distribution defines those zones, the data of which would suffice for adequate planning of a public transport network, for example. It has always been clear that for a project of this sort it is sufficient to rely on the data relating to zones of a high weight of importance: The road network based on them will cover the needs of the other, less active zones; and much planning time and expense will be saved.

In the framework of the Transportation Master Plan for Tel-Aviv, the public transport network indeed was based primarily on dominant zones, defined as such on the basis of the trip-ends distribution. When travel desires were assigned to the road network, the bus lines thus planned met about 90 percent of the direct travel desired in the city, with no transfers necessary. Thus, the ability to define the dominant zones in transportation distributions is, perhaps, one of the most practical results of the application of the KD.

Changes in Rank Size of Zones

Comparisons of the straight lines between periods of time might show that some interchangeability in rank size is taking place between certain zones: Some of the dominant zones are liable to relinquish their status with the passage of time, and vice versa. By considering these changes, planners might better allocate the limited economic means to the most important ends. In zones expected to lose their dominance, invest-

ments in transportation devices should be restricted to short term ones. On the other hand, in zones apt to become dominant, long term investments should be considered, even though present conditions do not warrant such investments.

Comparisons of Characteristics

Another noteworthy aspect is the possibility of comparing the characteristics of different K distributions. The significance of comparing the slopes of the lines in different distributions already has been pointed out. Another useful comparison is that between the dominant zones of different distributions. Taking again the trip-ends distribution, for example, the dominant zones in producing trips can be compared to the dominant zones attracting them. If comparisons reveal that the two groups consist of entirely different zones, then it should be clear to the transportation engineer that he is confronted with a pattern of widely dispersed trips. The various degrees of identity between the dominant zones of origin and the zones of destination, defined by such comparison, may provide the planner with an immediate, clear picture of the travel pattern he is confronted with.

Evaluations of Surveys

Finally, the reliability of the results of various surveys or projections may be established by examining the existence of the KD characteristics. If it is known, for instance, that the distribution of land uses in a certain town is highly concentrated whereas their attracted/generated trip-ends show a dispersed pattern, then a warning is due regarding the accuracy of the findings. The same warning applies to a specific zone that is dominant in the land-use distribution but falls below the breaking point in the trip-end distribution. It is to the planner's advantage to roughly evaluate the reliability of the findings easily and quickly at a very early stage.

CONCLUSION

Studying the various applications of the KD to transportation issues leads to the following conclusion. The technique of fitting the KD to transportation functions is a single device of manifold uses that enables immediate clarification of the following important basic concepts: the level of traffic urbanization and its trends of variation over different periods of time, over different modes of travel and over different cities; the regularity of development and the interrelationship between the traffic zones that constitute the planning units in transportation; the range of dominant zones that might serve as the basis for some major transportation projects; and, finally, the trends of variations in the dominance of traffic zones and their significances.

It should be noted, nevertheless, that because the KD is primarily empirical, its application is not intended to produce clear-cut solutions to problems that cannot be solved otherwise. Its application, however, can ease immensely the work of the transportation planner by providing an additional tool for analysis.

The initial experience of applying the K distributions to certain practical aspects in the preparation of the Transportation Master Plan for the Tel-Aviv area enabled us to confirm that these outlined advantages of its use have been fully attained.

REFERENCES

1. United Nations Demographic Yearbook, 1962, pp. 383-386.
2. Chicago Area Transportation Study, Vol. 1, p. 115.
3. Pittsburgh Area Transportation Study, Vol. 1, p. 104.
4. Tel-Aviv-Yafo Transportation Master Plan, Vols. 1 and 2. Prepared for the Municipality of Tel-Aviv-Yafo, Ministry of Transport, Dan and Egged Bus Cooperatives, by Kolin and Zahavi, Transportation and Traffic Engineers.

An Operational Planning Information System for Small Communities

ROBERT L. MAXMAN, Montgomery-Greene County Transportation and Development Planning Program, Dayton, Ohio; and
WILLIAM L. GRECCO, School of Civil Engineering, Purdue University

The wealth of data collected on the urban area by many people for a multiplicity of purposes has led to an inefficient, disorganized utilization of resources for data handling. Until recently, most of the information collected has been gathered by a specific group for a specific purpose. This information was not usable by other than the primary data recipient because of its narrow definitions and specific characteristics. Provided herein is a system whereby data that are collected only once are usable by all segments of the urban environment. Universally compatible definitions, aggregation units, and procedures are described. Computer programs were developed to handle the data for the system. The Environmental Data Storage and Retrieval System (EDSARS) will make a useful tool for all segments of the urban environment by putting all generally usable data in one place with one set of definitions and aggregated on one useful module, using one set of data handling procedures. The basic unit of data collection established was the parcel, thereby providing a high degree of flexibility in data aggregation. The conceptual development of information theory as it applies to urban data systems is first explored. The actual conceptual development of EDSARS is explained next, followed by the operational procedures needed to utilize the EDSARS system.

• FUNDAMENTAL to the planning process is the development of alternative plans, which are given to the decision-maker for action. These plans develop as a result of thorough analysis of proper and sufficient data. Analyses are good as, but no better than, the data quality. The present trend seems to be to seek more symptomatic relationships, which implies more data variables. Today it appears almost natural for researchers to add variables to the correlation analysis to increase the amount of variability of the dependent variable that can be explained. In spite of the general knowledge of the costliness of data collection, more data are being collected by more people for more reasons than ever before.

The value of data for proper analysis is not being questioned. The critical point is the efficiency of the entire data system, not from the standpoint of an individual user but in terms of total system costs. Data are being collected on many aspects, from the individual's health to the number of trips he makes. In the past, most of this information was gathered by a specific agency for a specific use, each agency applying its own individualized definitions to the data. For example, definitions of land-use density range from trips generated per acre to persons per square foot of floor space. The definition always has depended upon the information user. This multiplicity of data defini-

Paper sponsored by Department of Urban Transportation Planning.

tions and uses results in a hodgepodge of data collected many times by many agencies, often without knowledge of each other's efforts (1).

Public agencies gather much of their data for normal operations; these data could be very useful to other agencies at little additional cost if only common definitions and parameters could be established. Without these common definitions and parameters, it is difficult for each agency to visualize the urban environment except in the perspective of its own narrowly defined information requirements.

Each urban area has developed a multiplicity of plans to channel its growth in a manner that is deemed best for the community as a whole. Just as the different agencies collect and use their own data, so do the various urban studies. Whereas a recreational or school plan may use much of the data collected in a transportation study, the differences in data definitions and aggregation units make the information nearly useless for any study other than the one for which the data were collected.

There are many segments of the urban environment that desire information about the community. Any company, organization, or group that presently wants data on the community must collect the information itself or accept the narrow definitions of the data now established by existing governmental groups or studies.

To facilitate the use of data by other than the primary information receiver, a set of universally compatible definitions is needed. This set of definitions is not impossible to develop if one attempts to direct the collection of "pure" data. The term "pure" simply means that the information should not be aggregated before collection. For instance, when square feet of space is collected, it should be recorded as square feet, not square feet per some other dimension. Information concerning square feet per employee may be useful to an industry, but a knowledge of square feet of building and the number of employees is much more useful for planning purposes and still serves the original purpose.

The data system that is described here is an attempt to develop a tool for urban decision-making that utilizes data from many sources and makes this information available to and usable by other sectors of the urban community.

THE SCOPE

The project described involved the development of an urban data system for an area of approximately 100,000 population. The Lafayette-West Lafayette area in Indiana was used to demonstrate the application of the developed system. In the Greater Lafayette area, the conduct of a land use and transportation study was under consideration. In anticipation of the development of these data, this planning information system was developed. The system, referred to as the Environmental Data Storage and Retrieval System (EDSARS), was to be much broader than the proposed study (2).

The first developmental problem was choosing the degree of sophistication needed for such a system. This involved choosing a particular level in a hierarchy of data system complexity. After the level of sophistication had been decided, the basic data collection and aggregation module was chosen. The data to be used were selected along with specific definitions for each item. The methodology for entering these data into the system was developed along with updating procedures to keep information current. A logical and easily used means of data storage and retrieval was developed to facilitate use of the system by a wide variety of people.

DESCRIPTION OF EDSARS

Level of Data Sophistication

The information used in EDSARS is taken from the data library level in data hierarchy. Banked data, another level in the information hierarchy, are organized into machine records but need not be functional or logical in format. Raw data make up the lowest level of data sophistication. These data are not machine digestable and therefore are not usable in an organized data system. The data library information is logical and functional in format and can be updated, searched, and retrieved; these requirements are essential for any urban data system.

Level of System Sophistication

The three levels of system sophistication vary in the complexity of models incorporated. The first level uses no models, the second uses specialized models, and the third uses simulation. EDSARS, which is an attempt to develop the initial phase of an urban data system, uses the first level of sophistication. The system contains tabulated data but no specialized or simulation models. It is felt that the model requirements will evolve from the users' demands on the data system. The addition of models to the system can be made within the present format; the data in the system will feed directly any models developed in the future.

The computer hardware that is incorporated also influences the level of system sophistication. EDSARS uses the CDC 6500 computer at Purdue University. The CDC 6500 is a general purpose computer, and the programming language is CHIPPEWA FORTRAN. The data system can be initialized and information retrieved or updated by merely submitting the correct program deck to the computer science center. The updating, retrieval, or initialization will be run just as any other job that is submitted to the computer. The information for EDSARS now is stored on tape. As the system is initialized and the amount of stored information grows, the incorporation of a "disk pack" will become feasible. A disk pack is a mountable disk storage device that enables random access of information. This direct access feature will save valuable computer time when the system searches large quantities of data.

The decisions on the level of system sophistication were the result of many factors. Models were not incorporated into the system because of the need for actual data to test the validity of a model. This project outlines the initialization of EDSARS without actually inserting real data. The amount of data needed to initialize the system makes initialization another entire project of at least 1 to 2 years in duration. Once the initialization is complete, the addition of models can be considered.

The decision to use the CDC 6500 computer was made in light of the hardware available. Purdue University now has an IBM 7094 computer that could handle a data system such as EDSARS. The 7094, however, is a second generation computer; this type of computer is now in the process of being phased out by many organizations, being replaced by a third generation computer such as the CDC 6500. Any work done in the future on data systems most probably will be done on the more advanced equipment such as the CDC 6500. The use of CHIPPEWA FORTRAN was the result of the authors' knowledge of the language and the efficient data-handling capabilities of the FORTRAN developed for the CDC system.

Data Module

The data module for EDSARS is the parcel. This aggregation module seems to be almost the universal choice of existing urban data systems. The parcel provides a flexible, multipurpose base from which to work. The data to be incorporated into an urban data system are easily keyed to the parcel. The tagging methods work well with the parcel module. The parcel forms a very useful aggregation unit in that it is the largest common denominator that can be used to build zones. Any zone in an urban area can be represented fairly accurately by a composite of parcels. This capability gives the system maximum flexibility in the designation of zones with a minimum number of data units.

The parcel in EDSARS is defined as all contiguous land under one ownership and one general land use. This definition closely parallels the parcel used in assessors' records. If two adjacent pieces of land are owned by the same person and used for the same purpose, they are listed as one parcel. If two adjacent parcels have different uses, they are listed as two parcels. This definition, being general, allows a certain measure of ambiguity in the designation of a parcel; the system has the ability, however, to join two or more parcels into one new parcel or to break one parcel into two or more parcels. This capability of redefining parcels allows the system to establish its own equilibrium as the data are used and reevaluated.

A special definition of the parcel is used when rights-of-way are coded. Each street segment and utility right-of-way is coded just as any other parcel. A street or right-

of-way is broken into block-long segments if the block length is 500 feet or less; if the block length is longer than 500 feet, the block is broken into segments of 500 feet or less. An intersection is taken as a street parcel. The parcel boundaries are defined as the right-of-way lines for the street segments. An example of an area divided into parcels is shown in Figure 1.

Data Tagging Methods

EDSARS uses both the name and location methods of tagging data. The name tag is the street address of a building or empty parcel. The street number, name, and type (e.g., drive, street, lane, etc.) all are noted in the name tag of the parcel or building. For rural areas the street number is replaced by the rural route number, and the street name is replaced by "Rural Route." The name method of tagging gives the system the capability of locating data for the user on a basis that is familiar to all segments of the urban environment. Street addresses are universally known and understood and, therefore, enable all potential users to be familiar with at least one retrieval method.

Street segments are coded by the street name and the number (in hundreds) of most of the houses on the street segment. A segment along a street called Main Street in which house numbers go from 100 to 225 would be coded as 100B Main Street, which means the 100 block of Main Street. This code gives the benefits of the name tag to street segments as well as individual parcels and buildings.

The location tag utilized by EDSARS is a rectangular grid coordinate system, which is superimposed over the entire development area. The grid coordinate uses 1 foot as the basic unit. The parcels and street lengths are tagged by the coordinates of their approximate centroid. The actual digitizing of the coordinates is accomplished by an automatic coordinate digitizer. By the use of a location tag, internal logic is added to the data in EDSARS. The coordinates facilitate the retrieval of data on an areal basis. Data for certain geographical segments of the development area can be retrieved directly with the use of the coordinates, and density computations become immediately possible.

Rectangular grid coordinates provide another very useful capability. A zone, such as a census tract or transportation zone, can be represented by the grid coordinates of its boundary. This is accomplished by representing the zone by a series of triangles and digitizing the coordinates of the vertices. By representing zones in this way, a dictionary of zone names and grid coordinates is developed. When any information is desired on a zonal basis, the coordinates of the zone are read, and each parcel is tested to establish whether it lies within the zone in question. The information for each parcel within the zone is retrieved and aggregated, thereby giving information on the desired zonal basis. Figure 2 shows a zone broken into triangles for coding.

To coordinate the actual data incorporated into the system and the tags for each parcel, a dictionary with the parcel number, building number, and street address (or block number for street segments) is developed. Another file coordinating each parcel number and grid coordinate then is initialized. The actual data are stored in conjunction with a parcel number. The data are related to the name and location tags through the parcel number-building number-street address dictionary and the parcel number-grid coordinate dictionary. The parcel number is merely a unique number of one to six digits given to each parcel. The numbers need not be consecutive or have any

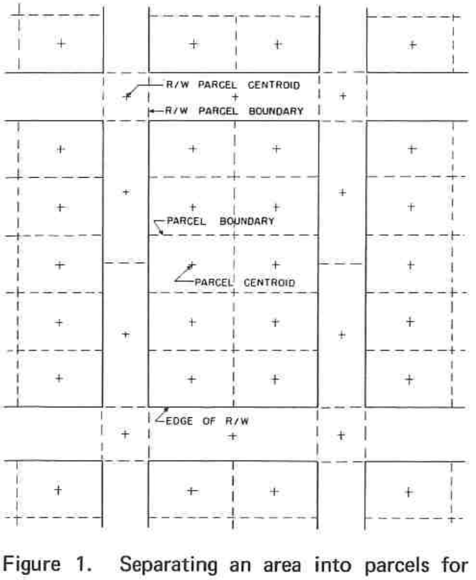


Figure 1. Separating an area into parcels for data coding.

logical order. The only requirement is that each parcel have only one number.

Data Dimensions

The definition of land use developed by the Metropolitan Washington Council of Governments was used in determining the data needed to define the different areas of land use. Data were examined in the light of how well they defined the following (3):

1. Type of activity
2. Type of structure
3. Type of land use
4. Intensity of use
5. Aesthetic qualities
6. Restrictions on use
7. Nuisance characteristics
8. Economic functions

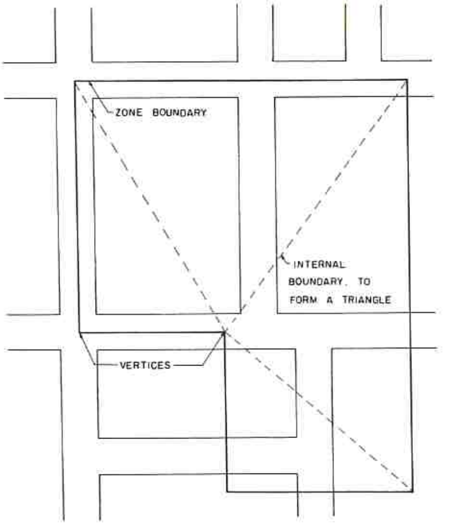


Figure 2. Zone divided into triangles for data coding.

To completely describe the urban environment, the information on each parcel is arranged into three categories.

1. Parcel information—information on the parcel itself, including dimensions, restrictions, zoning, and use.
2. Building information—information on each building on a parcel, including age, value, type of construction, condition, and size.
3. Establishment information—specific information on each unit within a building such as a business, a dwelling unit, etc., including space use, number of employees, number of residents, age of residents, and number of vehicles.

Information relating to these categories is collected by local studies and surveys such as those conducted by the Louisville Metropolitan Comprehensive Transportation and Development Program (4 through 18). The categories contain the following data items:

Parcel information

1. Land use
2. Ownership
3. Frontage
4. Area
5. Year of subdivision
6. Assessed value of land
7. Easement
8. Landmark
9. Neighborhood characteristics
10. Land appearance
11. Number of structures
12. Year of zoning change
13. Zoning
14. Zone change request number
15. Variance number
16. Comprehensive plan use
17. Utilities
18. Parking spaces
19. Loading area
20. Assessed value of improvements
21. Total assessed value
22. Sale date
23. Sale price
24. Nuisances

(The following data are collected for street segment parcels.)

25. Intersection
26. Length of segment
27. Right-of-way width
28. Pavement width
29. Functional class
30. Structural composition
31. Percent grade
32. Average daily traffic
33. Number of accidents
34. Traffic control signs and signals
35. Speed limit

36. Curb parking regulations
37. Curb type
38. Sidewalks
39. Number of lanes
40. Loading zone
41. Bus route
42. School route
43. Access control
44. Condition

Building information

1. Year built
2. Type of construction
3. Type of structure
4. Building condition
5. Year of latest building permit
6. Cumulative cost of building permits
7. Number of floors
8. Total floor area
9. Basement
10. First floor area
11. Number of dwelling units
12. Building setback
13. Required setback
14. Rehabilitation cost
15. Type of building code violations
16. Number of building code violations
17. Number of establishments

Establishment information

1. Space use
2. Total number of employees
3. First floor area
4. Total floor areas
5. Number of rooms for rent
6. Number of residents by sex and age group
7. Family income
8. Vehicles owned
9. Police calls
10. Fire calls
11. Welfare payment
12. Number of communicable diseases
13. Type of communicable diseases
14. Rent

Each data item that was entered into the system was judged to be important to the planning community, able to be updated, and relatively easy to collect. Data that were too expensive to collect or not updatable were not incorporated into the system.

OPERATION OF EDSARS

Data File Characteristics

The data in EDSARS are in four files. The first data file contains parcel numbers and parcel grid coordinates. Each parcel is given a unique number to identify it as a data entity; this number is correlated to the grid coordinates of the approximate parcel centroid by the parcel number-grid coordinate file. The second data file contains the parcel number and building number and street address of each building in the system. In the case of a parcel with no building, the building number is listed as zero. Each building is given a unique number on a parcel and coordinated to the correct address by the parcel number-building number-street address file. The third data file contains the parcel number and all general data on that parcel. The data for each parcel are correlated to the street address and grid coordinate via the parcel number. The fourth data file contains zone names and the coordinates of the zone boundary.

These four data files make up the data storage portion of EDSARS. The actual data are stored on tape and can be manipulated by a set of package programs. The first set of programs initializes the system by reading cards and writing the information on tape. The second set of programs incorporates more data items when they become available. The third set of programs reads the data tape and writes on paper. This set of programs checks the other program sets and gives a complete list of all data in the system. The fourth set of programs updates the values of data items in the system when current information becomes available. The fifth and final set of programs retrieves information

from the system for special user purposes. The following is an explanation of the package programs and procedures for using them in EDSARS.

General Program Information

Each of the programs discussed below manipulates the data in EDSARS. When data cards are read into the system, a card with "7, 8, and 9" punched in column one must follow the main body of the program and precede the actual data cards. When general data cards are read the last card should have a "99" in columns seven and eight. This indicates end-of-data to the system. The last data card in the zone-grid coordinate file, the parcel number-building number-street address file; and the parcel number-grid coordinate file should be blank to indicate end-of-data to the system. The last data card is followed by a card with "6, 7, 8, and 9" punched in column one. This indicates end-of-program to the computer. If no data cards are used in a program, the "6, 7, 8, and 9" card immediately follows the main body of the program.

Data File Initialization

Parcel Number-Grid Coordinate File—The initialization of the parcel number-grid coordinate file occurs first. To initialize this file, a coordinate digitizer is used (2). An accurate base map of the entire development area is placed on the digitizer, and a point 1,000 feet south and 1,000 feet west of the southwestern corner of the development area is set as point 0,0. Key points on the base map are digitized. The digitizer gives readings in inches, and these readings are converted to feet; the conversion depends upon the scale of the map used.

The key points digitized are the intersections of street centerlines. The key points located on the base map are digitized on the more accurate maps, and the coordinates of these key points serve as reference coordinates for the digitizing of all parcels.

Each map to be digitized, other than the base map, is first broken into parcels and approximate centroids located as shown in Figure 1. Consecutive parcel numbers are written on the map for each parcel. The map is placed on the digitizer, and its key point located (for coordinate conversion to the base system); then each parcel centroid can be digitized. The digitizer will punch the parcel number and the grid coordinates on an IBM card that can be fed into the computer for coordinate conversion.

The data cards for actual initialization of this data file are read into the system via the "Initialize Parcel Number-Grid Coordinate File" program. The data cards for this file should have the format shown in Figure 3. This figure also can serve as a sample coding sheet. The parcel number is placed in columns one to six. The x coordinate is placed in columns eight to 13 and the y coordinate in columns 15 to 20. When more information becomes available, a program titled "Read More Information for the Parcel Number-Grid Coordinate File" is used. This data file locates all the parcels in the development area and coordinates the parcel location to a particular parcel number. This number is the identifying tag in the system used to locate all data that pertain to this particular parcel.

Parcel Number-Building Number-Address File—The parcel number-building number-address file is initialized by putting the parcel number in columns one to six and the building number in columns 20 and 21. The street or rural route number is placed in columns 22 to 27. If the parcel is a street segment, the block number is placed in columns 20 to 26, with a "B" in column 27 to indicate "street block." All these numbers should be right justified. Columns 30 to 69 contain the street name or "Rural Route." The name should start in column 30 and be punched on the card just as it appears in the town directory, with one column between each word in the name. The type of street is coded in columns 70 to 72. Format for this file card is shown in Figure 3. When more information becomes available, a program titled "Read More Information for the Parcel Number-Building Number-Address File" is used. This file coordinates each building, vacant parcel, and block segment with a particular address in the development area.

General Data File—The file for general data is initialized after each parcel in the development area has been given a unique number. The format for general data cards is shown in Figure 4; this figure also can serve as a sample coding sheet. The 01 card is

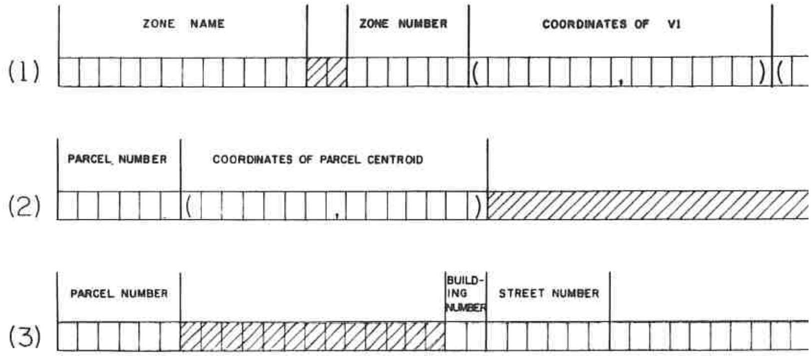


Figure 3. Format of (1) zone name-grid coordinate card; (2) parcel number-grid coordinate ca

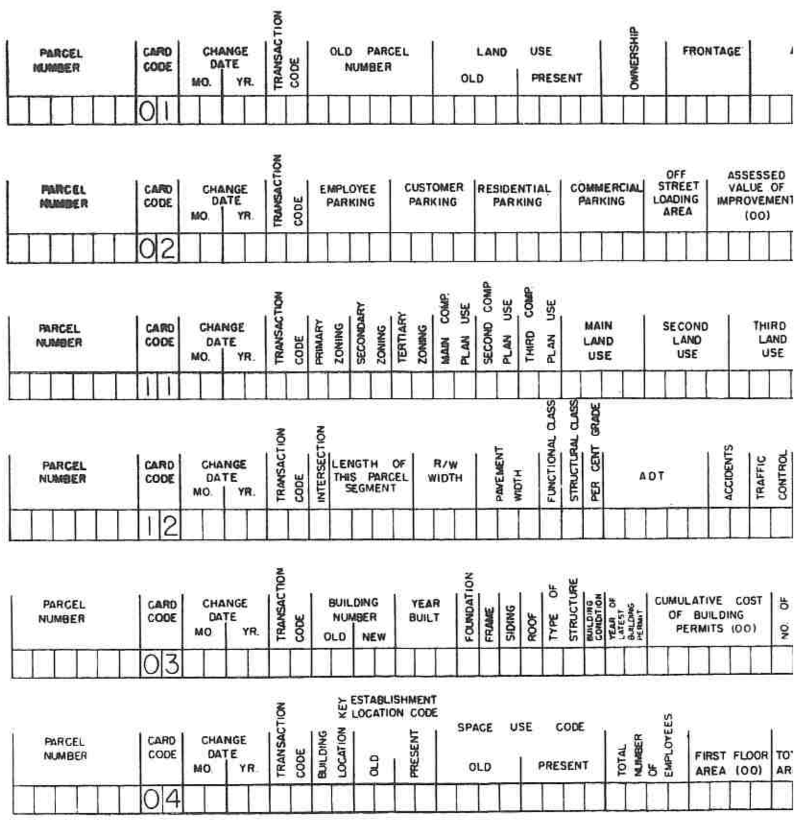
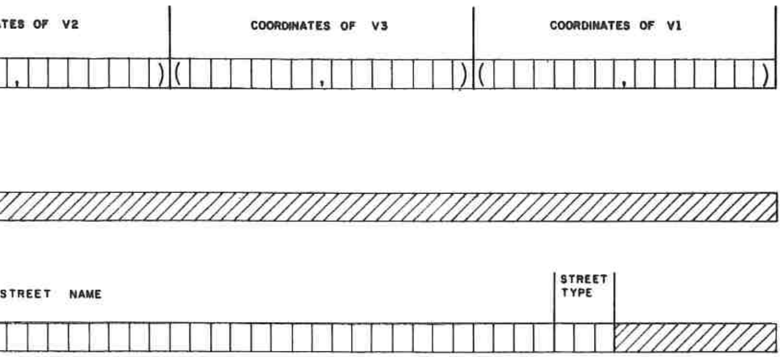
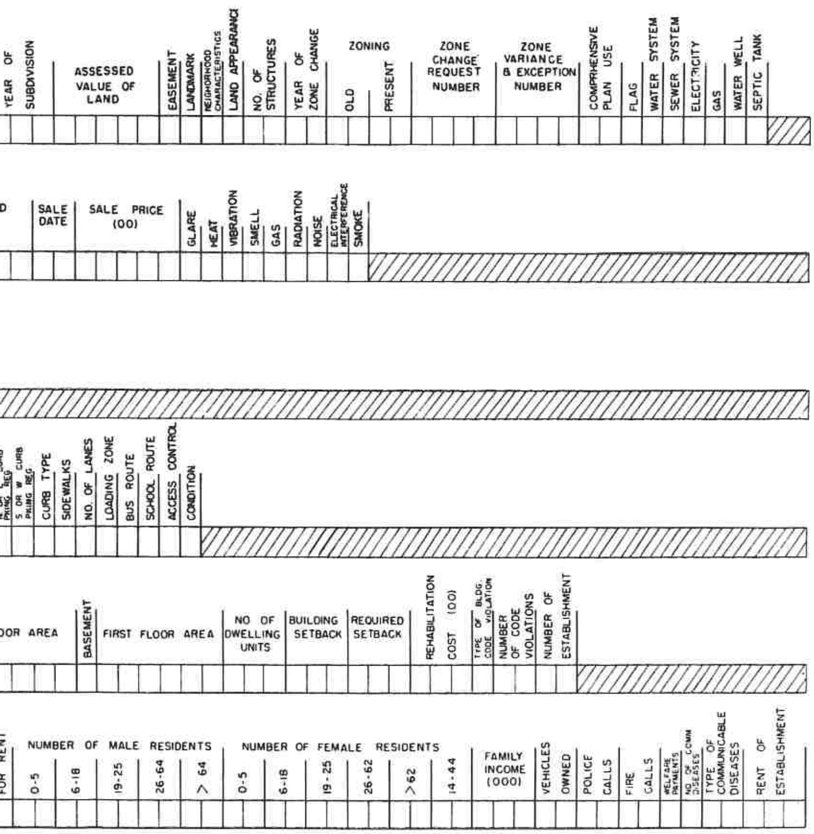


Figure 4. Format of general data cards.



parcel number-building number-address card.



used for every parcel in the development area. The 02 card is used when the parcel has a use other than right-of-way. A 12 card is used in place of the 02 card when a parcel has a right-of-way. If the parcel has multiple land uses or zoning or comprehensive land uses, an 11 card is used to supplement the 01 and 02 cards. Each building on a parcel is represented by the information on the 03 card, and each establishment (dwelling unit, business, office, etc.) in a building is represented by an 04 card. The information in the general data file is broken into three categories. The first category is land-use information and is represented by the 01, 02 or 12, and 11 cards; the second category is building information, which is represented on the 03 card; and the third category is establishment information, which is represented on the 04 card.

Each data item in the system is given a specific name that can be used to refer to the particular item.

Each building on a parcel is given a number to uniquely identify it; if four buildings exist on one parcel, they would be numbered one to four. Each establishment within a building is also given a unique number to identify it. Building numbers start at one in each separate parcel; establishment numbers start at one in each separate building. A program titled "Initialize the General Data File" initializes this data file by reading data cards and writing the information on tape and paper as a check. The program "Read More Information for the General Data File" reads in more information as it becomes available.

Zone-Grid Coordinate File—The last file to be initiated is the zone-grid coordinate file. To define a zone, its boundary is located in the development area by the grid coordinate system. The zone is broken into triangles, and the grid coordinates of each of the three vertices are coded on data cards. An example of a zone broken into triangles is shown in Figure 2. The card format is shown in Figure 3. The coordinates of the vertices are placed on the data card as follows. The vertices are numbered one to three; point one is coded first, followed by point two, point three, and point one again. The first and last coordinates must be the same in order to close the triangle. The identifying zone name is placed in columns one to 12 and the zone number in columns 15 to 20. The coordinates of the vertices are placed in columns 22 to 80 in the format shown in Figure 3. The zone name starts in column one. The zone number and grid coordinates are right-justified. The program "Initialize the Zone Name-Grid Coordinate File" initializes this file by reading data cards and writing the data on tape and then on paper as a check. To read more information into the system as it becomes available, the program "Read More Information for the Zone-Grid Coordinate File" is used. It should be noted that this file can contain as many zonal systems as required by the users. Census tracts, transportation zones, school zones, etc., are all examples of possible zonal systems that could be incorporated into this file. The inclusion of a particular zonal system is dependent upon the potential use of its parcel aggregation.

Read Programs

There is a general class of programs in EDSARS that will read the data file tape and print the information on paper. These programs should be used after reading in more data or updating the system to check the accuracy of the tape file. These programs also can be used to obtain a complete list of all information on the tape. These read programs will read the parcel number-grid coordinate file and print a complete list of the file; read the parcel number-building number-address file and print a complete list of the tape; read the general data file and print a complete list; and read the zone-grid coordinate file and print a complete list of this information.

Update Programs

To change or update any information in the system, a set of update programs has been developed to replace the old information by using the initializing programs to make a tape file of the new information. This new information file and the original file then are used to initialize a new tape file with all of the new information incorporated in it.

All the update programs require that the new data cards be identical to the original in format. The new cards should be complete—all information that is not changed still

should be punched on the update card. The new information is punched on the new card in the same format and these cards are used to form the update file. Any data card that has no data changed need not be entered into the update file, but any card that has any piece of information changed must be completely repunched with all the new and unchanged information.

Retention of Old Data Files

Data files represent current data for a certain period of time. The comparison of data files for different time periods can yield useful information on trends that exist in the development area. It is felt that files should be updated at least once a year. These yearly files should be retained for at least 5 years. The final decision on this policy is, of course, up to the initializing agency.

RETRIEVAL PROGRAMS

To retrieve information from the data files for special purposes, there are programs in EDSARS that give specific information for special purposes of the user. The following programs were designed to be general in their characteristics so that specific user needs could be satisfied. The programs available in the system are designed to retrieve

1. A list of y and x;
2. The sum of y for a specific x;
3. A list of y for a specific x;
4. x for specific parcel numbers;
5. A parcel number and building number for a specific address; and
6. A list of parcel numbers for a specific zone.

CONCLUSIONS

The following conclusions about EDSARS and its potential can be made:

1. EDSARS should facilitate efficient and economical handling of planning data for an area of about 100,000 population.
2. The utilization of a general purpose computer and general purpose programming languages should make EDSARS available to most metropolitan areas in the United States.
3. The concept of a unified data system is the most important contribution of EDSARS.
4. The data proposed for EDSARS are the most usable and easily obtainable information available to the urban area.
5. The incorporation of a flexible method of representing zones by their location is essential to an efficient urban data system such as EDSARS.
6. The information for an urban data system should be in four separate files so that one file can be updated and improved without disturbing the other files.
7. Zone names and boundary locations should comprise one file; parcel numbers and parcel location should comprise another file; parcel numbers, building numbers, and street address should comprise the third file; and the fourth file should be made up of general data.
8. The best unit for data collection is the parcel.
9. The data system should be flexible so that improvements can be made as the system is used and technology increases.
10. The streets and rights-of-way should be represented as special parcels to ensure full territorial and informational coverage.
11. All data incorporated should be potentially useful and updatable.
12. Utilization of applicable theory and practical experience of existing data systems is needed to develop a useful, efficient, and improved data system.

The concepts represented by these conclusions, tied together in an urban data system such as EDSARS, give the planning community and the urban environment as a whole a flexible and useful tool. The EDSARS system should make more information available to more people at a much lower cost and with much less effort.

ACKNOWLEDGMENTS

It is important to acknowledge that the conceptual technique for retrieving information on a zonal basis through the use of coordinates was developed at the University of Washington by Mr. Robert B. Dial. The details of the EDSARS system, including initialization, update, and retrieval programs; codes; and complete descriptions are available in a report completed at Purdue University (2). Copies are available at the cost of reproduction. This project was funded by the Public Health Service through the Environmental Health Institute of Purdue University.

REFERENCES

1. Metropolitan Data Center Project. Tulsa, Oklahoma, February 1966.
2. Joint Highway Research Project Information Report 5. Purdue University, Lafayette, Ind., March 1968.
3. Interim Report: Metropolitan Planning Data From Local Governments. Metropolitan Washington Council of Governments, August 1966.
4. Home Interview Manual. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
5. Truck and Taxi Manual. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
6. External Survey Manual. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
7. Land Use Manual. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
8. Road and Street Inventory Manual. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
9. Flow Characteristics Manual. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
10. Accident Study Manual. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
11. Parking Study Manual. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
12. Routes and Terminals Manual. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
13. Transportation Facilities Standards. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
14. Population Economic Base Manual. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
15. Planned Transportation Facilities. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
16. Public Transit Study. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
17. Development Data. Louisville Metropolitan Comprehensive Transportation and Development Program, 1965.
18. Maxman, Robert J. The Data Flow Collection to Model Calibration for the Louisville Transportation Plan and Its Application to Data Analysis. Louisville Metropolitan Comprehensive Transportation and Development Program, Vol. 1, No. 1, pp. 1-27.

A Methodology for Forecasting Peak and Off-Peak Travel Volumes

MARTIN WOHL, The Urban Institute

No information can be more important to transport planners, designers, and analysts than reliable forecasts of the peak and off-peak travel volumes on transport networks. Yet, no reasonably complete and valid methodology has been proposed—much less developed and verified—that will permit the transport planner and analyst to properly differentiate travel by time of day; to determine realistically the duration and level of peaking and recognize its dependence on the transport system design and performance; and to account for shifts in trip-making from one hour to another, from one mode to another, and from car pooling to driving alone in response to changes in transport system design, in service, or in price. Further, the usual travel forecasting process treats trip generation as though trip-making were independent of transport system changes, and it treats the trip distribution, modal-split, and route assignment phases as though trip-making choices—whether to travel, final destination, mode, and route—are made sequentially and apart from the circumstances attendant with the other phases. Accordingly, the purpose of this paper is to formulate a model structure such that these phases can be treated simultaneously, that the total amount of trip-making (as well as the destination, modal, and routes choices) can be varied with the transport system and its performance and price characteristics (among other factors), that shifts from car pooling to driving alone can be represented, that shifts from one hour of travel to another can be characterized, and that the amount of travel during peak and off-peak hours (i.e., the absolute buildup or decrease in peak or off-peak flow) can be determined.

•THE TRANSPORT PLANNER and analyst have fashioned numerous models dealing completely or in part with the travel forecasting problem. The literature is vast and hardly needs repeating or a review. To my knowledge, though, none of the available models and techniques deal realistically or structurally with the matter of peaking.

More precisely, none of the models and techniques appropriately recognize that travel during peak periods is dependent both in concept and in actuality on the nature and extent of the transport system, on its performance and price characteristics during both peak and off-peak periods, and on the attendant socioeconomic conditions and preference patterns of travelers for peak and off-peak travel choices.

As we design, analyze, and evaluate marginal or major adaptations to an existing system, what is it that we need to know? In part, we want to know how usage and performance of the system and its parts will be affected. How much extra travel will there be? Which and how many people will shift from one road or mode to another? How many people will shift out of car pools either to driving alone or another mode? Obtaining data by time of day is vital. Will the peak-period volumes increase? Will people shift from other hours of the day to the peak period as capacity is added? Will the peak period shorten and by how much? Will total daily traveling increase?

Paper sponsored by Department of Urban Transportation Planning and presented at the 49th Annual Meeting.

Such questions cannot be answered by our current models and certainly not by either our simplistic peak-hour factoring techniques or our trip-purpose models developed as a proxy for time-of-day or peaking models. Nor can the analyst properly evaluate the benefits and costs stemming from one design or another without having knowledge of both peak and off-peak travel conditions.

To formulate a methodology suitable for forecasting travel volumes and performance conditions during peak and off-peak periods, distinctions must be made between the prediction process requisite for shorter time periods and that relating to longer time periods. For the former, the transport system, population, employment, and land-use patterns can be regarded as fixed. The latter considers the transport system as it affects and is affected by the land-use pattern as well as the growth and distribution of population and employment. It also will be necessary to distinguish between "demand" and "supply." Demand involves the propensity of people to travel with respect to travel service, price, and socioeconomic conditions. Supply describes the performance of the transport system with respect to the amount and composition of travel sustained by it.

LONG TERM VERSUS SHORT TERM FORECASTING

For the distinction pertaining to the time frame for our forecasting, we must ask: Are we attempting to determine the amount of travel taking place at some point in time, given the transport system, land-use, population, and employment patterns? Or are we trying to develop a more dynamic forecasting model capable of forecasting both the short- and long-range travel and land-use conditions?

For the first of these time-frame questions, we must be concerned with travel forecasting in some static or partial equilibrium sense. Specifically, as a basis for analyzing and evaluating our planning, policy, or design actions, we need to know how much travel will take place and the associated travel conditions, both hourly and daily, given the following information:

1. The socioeconomic characteristics of the people;
2. The location and character of business, industry, and residence; and
3. The physical and operating characteristics of the transportation system.

For such short-run or daily travel forecasting involving the transport system, home and business locations and the transit fleet can be regarded as fixed. By contrast, it is hardly clear that the automobile fleet or ownership should be regarded as fixed, even when forecasting trip-making and modal-split over the short run. As travelers choose among modes on a day-to-day basis, many or most of those who travel by auto, particularly those driving alone, probably made that modal choice at the time they purchased the auto and thus are not making a new decision based on the marginal daily service and price circumstances each day.

This is not a simple problem conceptually or operationally, but it is an important one. In terms of predicting the number or percentage of travelers using one mode or another, this auto distinction may not seem important because most auto travelers are car poolers and may well be viewing the travel conditions for the various modes based on the marginal day-to-day circumstances. In terms of examining traffic congestion, however, and the effects of changes in mode or capacity on its reduction, it is the number of drive-alone vehicles that is most important because these vehicles represent the great bulk of the total auto fleet during peak hours and because their drivers probably made their modal-choice decision on more than day-to-day marginal costs.

Also, and to cast this matter in a slightly different fashion, consider the urban dweller who is examining the tradeoffs associated with a suburban versus central-city dwelling unit. Although the living space, privacy, school conditions, type of neighbor, and housing cost are probably the most important factors he considers, no doubt he also takes into account the available modal choices with respect to travel time, convenience, and cost. For the latter, he probably thinks about the total auto ownership and operating costs because a second car often will be required. In short, he buys the second car based on the day-to-day travel time and convenience expectations and on the long-range travel cost factors. If these hypotheses are correct, our modal-split models must indicate these short- and long-range considerations.

Once the analyst has developed the capability of forecasting daily travel volumes and performance conditions, he can turn to the more formidable problem of long-range forecasting. He can ask a host of location-living-transportation behavioral questions such as: How does the buildup of congestion and the attendant costs, taken with other factors of production and preferences (with respect to patterns of living, quality of life indices, work/employment/shopping/business locations, and so forth), influence changes in employer's plant or business growth and location, and in home or work location? How do these shifts then change the performance of the transport system, which in circular fashion then influences other locational shifts or growth patterns? How will native preferences about transportation services and living patterns (to take but two aspects of importance) change over time, either in response to income changes or in response to shifts in society's scale of values and mores? Because shifts in location and growth stem partially from expectations about the daily travel conditions at different points in time, a relation exists between long- and short-range forecasting. Essentially, the long-run changes in growth, location, and transport service are the result of the accumulated short-run or daily circumstances and conditions which occur over the longer time period.

The interrelationship and distinction between short- and long-range travel forecasting can be expressed in a number of ways, one of which is shown in Figure 1. This flow-chart representation for the general equilibrium or long term forecasting problem is particularly weak in at least one respect. Even though locational shifts and land-use growth do occur incrementally from year to year (or whatever time lag seems appropriate for modeling of this sort), one should not infer that the yearly shifts or growth result simply from the present-day equilibrium flows, prices, performance levels, costs, and so forth. Rather, it seems likely that dwellers and businessmen, in shifting to new home or work locations and making modal choices (decisions which are partially interdependent with the former), are responding both to the present-day transport, land-use, and socioeconomic conditions and to those that are expected for all (foreseeable) future years. As a consequence, the time-lag type of procedure for linking the static

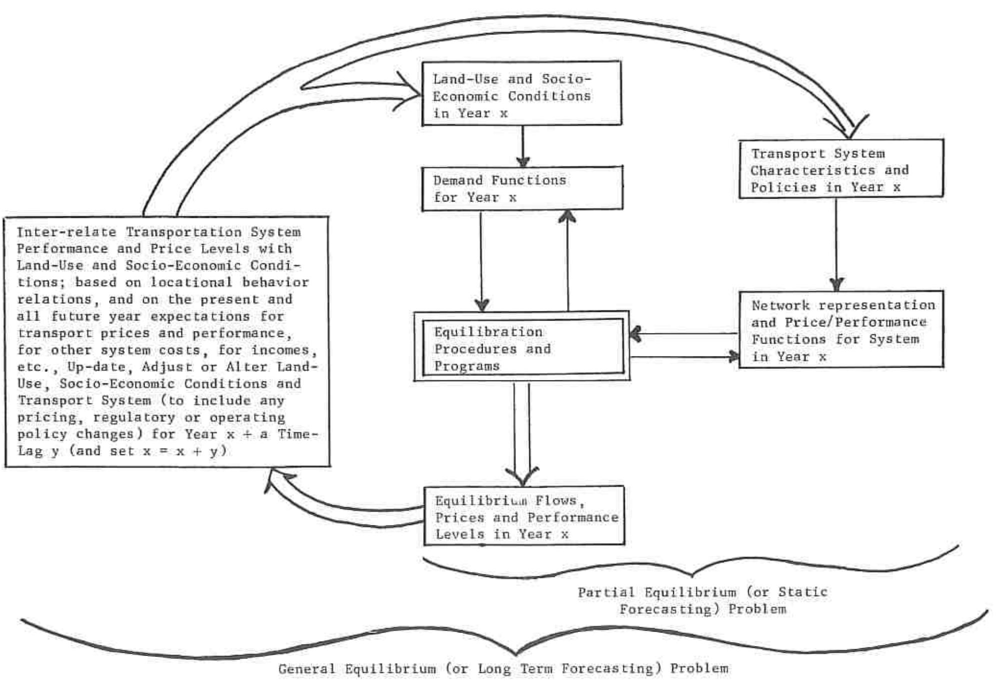


Figure 1. Schematic diagram for travel forecasting process.

and long term forecasting models is considerably more complicated than is illustrated here; the adjustments from year x to year x + 1 and some shortrun decisions rely not simply on the year x equilibrium conditions but on those for years x, x + 1, x + 2, ... N (assuming an N-year planning horizon).

THE ROLES OF DEMAND AND SUPPLY IN THE TRAVEL FORECASTING PROCESS

The distinction between demand and supply is an important one and is crucial to the formulation of an appropriate travel forecasting methodology.

Usually, demand is regarded incorrectly as the number of trips that will be made within, for example, the urban region at some future date. Demand is regarded, therefore, as the need or the requirement that must be met. This view of demand is somewhat analogous to the concept implicit in the trip-generation phase of most current forecasting models. As such, all that remains to be determined is between which zonal pairs and by what mode and route these trips will be made.

Contrarily, demand should be viewed as a statement of people's trip-making propensities; that is, it should be viewed as a demand function or conditional trip-making relationship. Thus, a demand function represents the dependence of the demanded quantity of trip-making on the price of or service afforded by trip-making. Implied, of course, is that more trips (in the absolute and relative sense) will be demanded (or generated) if either the price is reduced or the service is increased, whereas increased traffic congestion will tend to reduce trip-making. (As will be discussed in the next section, the demand function characterizes much more than trip generation; the demand function simultaneously incorporates the trip distribution, modal-split, and perhaps even route assignment phases as well.) Clearly, though, this functional and behavioral view of demand should give one pause when thinking about the common notions of "meeting the demand," or "needs," or "requirements," or constant trip-generation rates.

Another aspect of demand pertains to changes or shifts in demand. Over the short run, demand will not shift or increase; and thus changes in the amount of trip-making that occur in response to price or service changes should be regarded as movements along the demand function (or demand schedule) rather than as increases in demand. By contrast, increases in demand or shifts of the demand function will stem from long-run changes in population, income, tastes, and so forth.

The concepts of supply and demand are useful mainly because of the analogies that can be drawn from microeconomic theory, particularly in terms of specifying the interaction between supply and demand and of determining equilibrium prices and quantities demanded. Although a direct analogy can be made between the economist's and the transport analyst's characterization of travel demand and between their equilibration of supply and demand, there is only an approximate analogy for supply when applied to transport networks or links of a network.

In microeconomic theory, the term "supply" refers to the supply schedule—the amount of a product that will be supplied by the industry at different price levels. It is the amount supplied collectively by all firms producing that same product or service while assuming marginal cost pricing. In the context of this paper, "supply" is meant to characterize either the dependent relationship between travel service and the usage on the travel facility or that between travel price where the combined money and nonmoney time, effort, and expense of travel are placed on a commensurate value scale and usage. Alternatively, these expressions may be viewed as performance or service functions and are entirely analogous to capacity-restraint functions that have often been used in travel forecasting processes.

Employing the concepts of demand and performance functions to forecast trip-making at some point in time and for a given land-use plan and transport system requires that we follow a three-step process:

1. Describe trip-making behavior; i.e., specify demand functions (rather than point estimates or projections) of the form q = f(price, service, socioeconomic characteristics), where q is the quantity of trip-making demanded for the price, service, and other specified conditions.

2. Describe system service or performance; i.e., specify service-performance functions of the form p = f(system capacity, technology, controls, operating and price policies, volume of usage), where p is price resulting from the volume, capacity, and other specified conditions.

3. Interrelate supply and demand; i.e., equilibrate demand and performance functions for the region and transport network in question, so that point estimates of actual or equilibrium volumes and service or price levels can be determined.

In other words, we must find the values of p_x and q_x that will satisfy the following constraints:

$$q_x = f\left(\underline{p}_x, \underline{SE}_x\right)$$

$$\underline{p}_x = f\left(q_x, C_x, OP_x, PP_x\right)$$

where $\underline{p}_x$ is a vector of the price or performance conditions occurring in year x; q_x is the flow occurring in year x; C_x is the system capacity in year x; OP_x is the operating or control policy in year x; PP_x is the pricing policy in year x; and $\underline{SE}_x$ is a vector of the socioeconomic conditions in year x. The resultant $\underline{p}_x$ and q_x values are the equilibrium prices (or service levels) and flows and thus are the forecast for that system, that pricing policy, that year, etc. Figure 2 shows this interaction and the resultant or equilibrium price and volume levels.

Simplistically, the equilibrium flows and prices for a facility before and after improvement (for a one-link facility) can be as shown in Figure 3. As noted before, the induced traffic or increase in equilibrium flow from V_A to V_B that stemmed from the improvement and reduction in congestion or price should be regarded as a movement along the demand function or as an increase in the quantity of trips demanded rather than as an increase or shift in demand.

Demand, however, can and usually does shift or increase over time as a result of increases in population, income, etc., and because of changes in taste that generally affect the equilibrium flows and prices as shown in Figure 4. Thus, yearly increases in flow will stem from shifts in demand. Each yearly increase is generally slightly less than that for the previous year because of the exponential nature of queueing delays and thus the price-volume or performance curve.

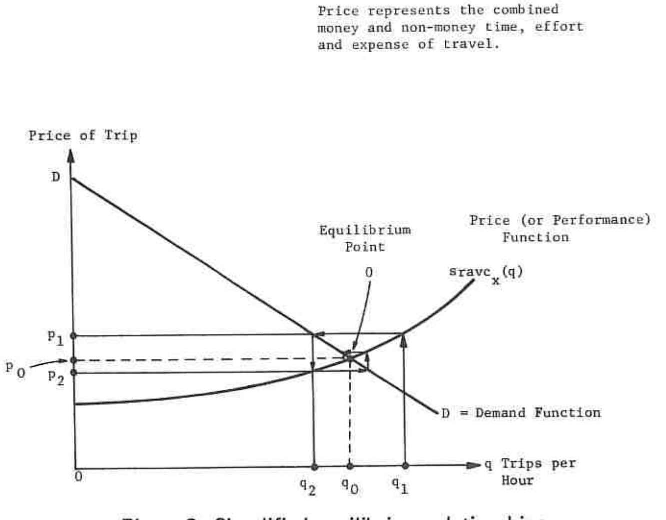


Figure 2. Simplified equilibrium relationships.

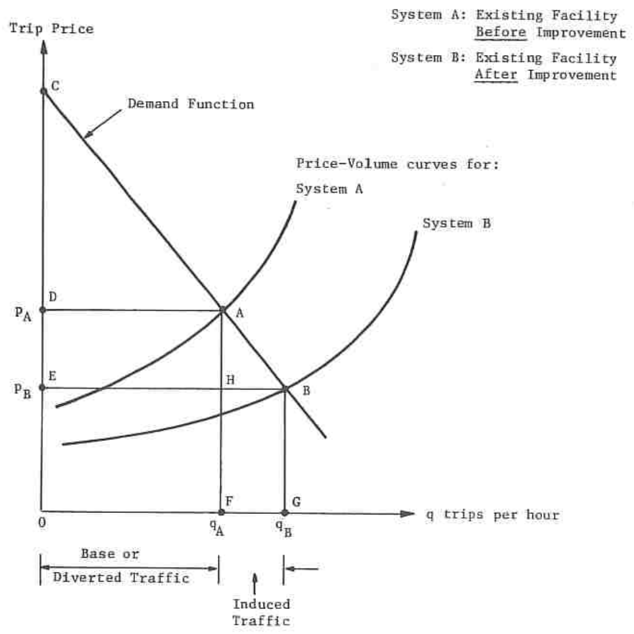


Figure 3. Equilibrium conditions for different facilities.

These simple notions about supply and demand and their interaction can be extended to explain hour-to-hour differences in trip-making and the peaking phenomenon that accompanies them. First, the demand for travel (as distinct from the equilibrium or actual flow) will fluctuate from hour to hour in response to people's preferences for traveling at specific times of day; in general, the demand for travel at starting-to-work or going-home-from-work times will be higher and less sensitive to congestion than that for other times of day. For illustrative purposes, then, demand throughout the day

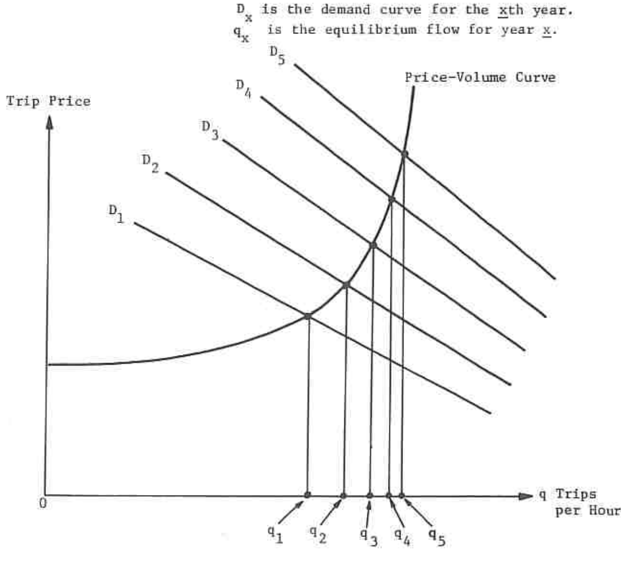


Figure 4. Intertemporal demand and price-volume relationships.

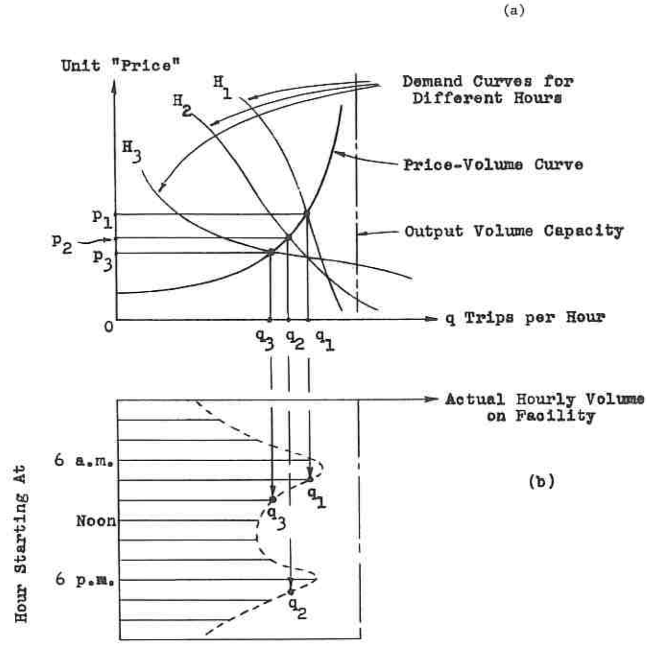


Figure 5. Short term intratemporal demand and price-volume relationships. For simplicity, demand curves for only 3 hours are shown. Also, for demand functions, the dependent and independent variable axes have been reversed. H_x is the demand for travel during hour x.

may be represented by a series of hour-by-hour demand functions as shown in Figure 5(a). (These relationships are oversimplified in some important ways which will be clarified in a later section.) Second, the interaction of the hour-by-hour demand functions with the price-volume curve determines the hour-to-hour equilibrium volumes which then may be plotted as volume versus time of day, as shown in Figure 5(b).

Considerably more discussion about the equilibration problem and about the full-scale development of appropriate demand and performance functions is given in the following section. These opening remarks are intended merely to introduce the aspects of forecasting and to indicate how demand and performance functions are related in the overall forecasting process.

DEVELOPMENT OF DEMAND FUNCTIONS

Demand When Considered in a Behavioral Context

The following points will be the basis of the model development.

First, demand for travel is a derived demand; that is, it is derived from a fundamental desire to do something else rather than travel without purpose. Thus, an understanding of travel relations must rely, to some degree, on the commodities and services being acquired at the trip destination. More simply, the value of a trip and, therefore, the extent to which it will be demanded depend on the importance of that trip to that individual. Is it a work trip? A pleasure trip? A doctor's visit? Is it important? Is the trip of little consequence and can it be foregone easily? Even work trips are given up when travel conditions are bad enough. These remarks suggest, at a minimum, that demand should be stratified by trip purpose.

Second, an individual's demand for goods and services depends on his social characteristics: family size, tastes, upbringing, income, etc. Stratification by income at least seems important.

Third, destinations differ in terms of the services and goods offered or number of opportunities; or they may differ locationally, aside from travel; or there may be just "perceived differences." We will need to specify demand, therefore, in terms of specific destination.

Fourth, in deciding whether to travel and what mode to choose, trip-makers invariably consider the circumstances for both directions of the trip. One may not go downtown by commuter railroad, for example, if he cannot come back until after the last train leaves.

Fifth, trip-makers choose modes on the basis of service and price differences and their value scales as they perceive them. Because the analyst's differentiation by service and price is not sufficient to explain trip-making behavior, we must assume that some influence variables are overlooked or improperly measured. Along this line of reasoning, mode-specific stratification should treat drive-alone car and car pooling as two separate modes.

Sixth, travelers probably view the route selection problem in a fashion somewhat analogous to choosing modes, though it is conceivable that route switching occurs along the route as events or information along the way changes one's perception. Route stratification does, however, seem in order.

Seventh, the hour of day for both ends of a trip appears to be an important consideration. The time of travel and the mode chosen are independent neither of one's time-of-day preferences nor of the travel conditions during the preferred and other-than-preferred times of day.

Characteristics of Simplified Demand Functions and Some of Their Forms

One essential characteristic of demand function is sensitivity or elasticity. For example, the "elasticity of demand with respect to price" is a dimensionless measure of the degree to which travelers respond to price changes. Specifically, the elasticity (e_p or η_p) is defined as the percent change in quantity demanded that accompanies a 1 percent change in price; or:

$$e_p \text{ or } \eta_p = \frac{\text{relative change in quantity}}{\text{relative change in price}} = \frac{\partial q/q}{\partial p/p} \text{ or } \frac{\Delta q/q}{\Delta p/p}$$

The elasticities for two forms of single-variable demand functions are shown in Figure 6. For a linear demand model the elasticity varies over its entire range whereas for a nonlinear model of the hyperbolic form the elasticity is constant.

From the designer's point of view, the measure of elasticity permits determination of the changes in toll revenues and volume and thus road capacity (or toll booths) which stem from altering the toll structure. For a transit operator, changes in the number of buses needed and gross revenues can be calculated.

The practical usefulness of knowledge about elasticities, which is virtually unused in urban transport circles, cannot be overstated. If transit fares, for example, are

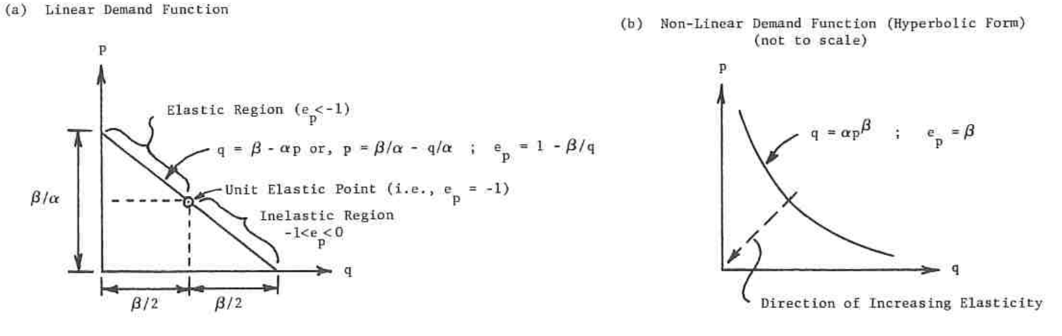


Figure 6. Single variable linear and nonlinear demand functions.

currently within the elastic portion of the demand curve, a decrease in price (without a change in service or schedule frequency) will increase ridership and gross revenues. (An increase in net revenues may or may not result, depending on the increase in costs stemming from the extra ridership.) On the other hand, if fares are presently within the inelastic region of the demand function, an increase in fare (without changing service or schedule frequency) will increase the gross and net revenues but decrease ridership. Similarly, the utility of such knowledge (that is, knowing if the demand is elastic or inelastic and to what extent) to toll authorities, railroads, airlines, etc., is all too obvious. Clearly, though, this knowledge can be exploited fully only by having similar types of information on the accompanying cost changes.

It is also important to recognize that utilization of this type of demand function, in contrast to the more usual trip-generation/trip-distribution approach, permits the analyst to assess directly the effect of price or performance changes. He is able to note increases or decreases in congestion caused by a change in technology or operating policy on the overall amount of trip-making or trip generation. Trip generation thus need not be calculated independently of or insensitive to the transport system performance characteristics and improvement. Furthermore, once these simplified single-variable demand models have been extended to incorporate multimodal and multi-time-period aspects, the ability to ascertain both the amount of trip-making by mode and by time of day and the shifts among modes and times of day can be achieved and the changes in trip-making, as well as in its modal and time-of-day distribution, that stem from changes to the transport system (or other influence variables) can be reflected.

Both of the previously mentioned linear and hyperbolic type of nonlinear demand models (Fig. 6), as well as a host of others, are in a form convenient for estimating parameter values. In both cases, linear regression techniques can be employed for estimating the parameter values (that is, for estimating α and β). For the nonlinear model, though, it is first necessary to convert the primary demand function into its log/log form. That is, given

$$q = \alpha p^{\beta} \tag{1}$$

and taking the logarithm of both sides of the equation, we get

$$\log q = \log \alpha + \beta \log p \tag{2}$$

which is a linear model (though in log form) that can be employed for estimating parameter values.

A third form that can be used for the demand model is the exponential in which

$$q = \alpha e^{\beta p} \tag{3}$$

Taking the natural log of both sides, we get

$$\ln q = \ln \alpha + \beta p \tag{4}$$

thus again providing a linear form for parameter estimation.

Extensions of Demand Models to Account for Modal Split and Peaking

First, demand models may be constructed to handle multiple service, price, or performance variables, such as enroute (or line-haul) travel time, access travel time, money price, and number of transfers. For example, if the money expense p and total trip time t were the only influence variables, the demand model might be formulated in one of the following ways:

$$q = \alpha - \beta p - \gamma t \tag{5}$$

or

$$q = \alpha p^{\beta} t^{\gamma} \tag{6}$$

Our interest then will focus on elasticity with respect to price and elasticity with respect to travel times. Quite properly, this more complete form (to include as many service variables as is appropriate) implies that the package of service-price levels is what influences trip-making. A priori, we would expect both the time and price elasticities to be nonpositive; in advance of gathering field information, however, nothing can be said about whether they fall within the elastic or inelastic region of the demand function or about whether price or time elasticities are largest. Knowledge about the elasticities in both absolute and relative terms is vital, however, because it will permit the analyst, designer, or operator to judge whether changes in service or price will beneficially affect ridership or revenues.

Second, and in a similar vein, these types of demand models may be extended to handle modal choice. One may differentiate between modes in one of two ways:

1. In terms of the technological difference, as usually is done, thus identifying bus, rail, and auto modes, for example; or
2. In terms of the service, price and performance differences, thus classifying modes only by differences in the service-price-performance package.

Should two technologies have identical service-price-performance characteristics, then travelers will be indifferent in choosing between them so long as all the price, service, and performance variables which influence the trip-making and modal choice have been incorporated. However, because travelers are not indifferent to the available modal choices but show a preference for one or another even though the price and service levels for the modes as defined and measured by the analyst are identical, then one may assume either that all the price-service-performance variables influencing behavior were not included or that the measurements of these variables were incorrect. In such a case, it will be necessary to abandon the latter type of differentiation and make use of the technological classification.

As a simple example, consider the demand functions for modes 1 and 2 with the price of each mode being the only measurable influence variable. Then, to take two alternative formulations,

1. Linear model form:

$$q_1 = \beta_1 - \alpha_1 p_1 - \gamma_1 p_2 \tag{7}$$

$$q_2 = \beta_2 - \alpha_2 p_1 - \gamma_2 p_2 \tag{8}$$

where $\alpha_1 p_1$ and $\gamma_2 p_2$ are direct demand relations, and $\alpha_2 p_1$ and $\gamma_1 p_2$ are the cross relations, which reflect the substitutability; and

2. Nonlinear model form:

$$q_1 = \alpha_1 p_1^{\beta_1} p_2^{\gamma_1} \tag{9}$$

$$q_2 = \alpha_2 p_1^{\beta_2} p_2^{\gamma_2} \tag{10}$$

For both model forms, the direct elasticities are respectively

$$e^1_{p_1} \text{ and } e^2_{p_2}, \text{ or } \frac{\partial q_1/q_1}{\partial p_1/p_1} \text{ and } \frac{\partial q_2/q_2}{\partial p_2/p_2}$$

The cross elasticities reflect the substitutability of one mode for another and are defined as the percentage change in quantity of travel demanded for one mode which accompanies a 1 percent change in the price of another mode; e.g., the cross elasticity of demand for mode 1 with respect to price of mode 2 is

$$e^1_{p_2} = \frac{\partial q_1/q_1}{\partial p_2/p_2}$$

It is important to emphasize that this type of demand model (in contrast to that implicit in the usual trip-generation/trip-distribution/model-split/route-assignment process) directly accounts for the following "real world" facts:

1. The travelers' decisions to travel or not and to select one mode or another are treated simultaneously. That is, one does not decide to travel irrespective of the alternatives afforded him and their service characteristics.
2. Both the amount of trip-making (summed over all modes) and the split among modes can and do vary with changes in travel service or price.

Third, but only to the extent that the amount or nature of trip-making is affected, it will be necessary for our demand models to incorporate the socioeconomic conditions of potential travelers who may originate trips at zone i and of the opportunities at a potential destination zone j. That is, $q(ij, m) = f$(transport service-price and socioeconomic variables), where $q(ij, m)$ is the quantity of travel going from zone i to j by mode m. Just one of many possible model forms might be

$$q(ij, m) = \alpha_m (Y_i)^{\beta_m} (P_i)^{\gamma_m} \left(E_j\right)^{\delta_m} \prod_{x=1}^{M} \left(p_{ij}^{x}\right)^{\theta_{m,x}} \prod_{x=1}^{M} \left(t_{ij}^{x}\right)^{\varphi_{m,x}} \tag{11}$$

where Y_i is an income measure for zone i travelers, P_i is a population measure for zone i, E_j is an employment measure for zone j, p_{ij}^{x} is the money price for trips from i to j by mode x, t_{ij}^{x} is the travel time for trips from i to j by mode x, and M is the number of travel modes available for trips from i to j.

For this hyperbolic form, as before, the exponents are the elasticities with respect to the particular variables; e.g., β_m is the elasticity of demand for travel by mode m with respect to the zone i income measure. Similarly, $\theta_{m,x}$ (for $x \neq m$) is the cross elasticity of demand for travel from i to j by mode m with respect to the money price for travel from i to j by mode x. The exponent $\theta_{m,m}$ is the direct elasticity of demand with respect to the money price for travel from i to j by mode m. Also, to clarify,

$$\prod_{x=1}^{M} \left(p_{ij}^{x}\right)^{\theta_{m,x}} = \left(p_{ij}^{1}\right)^{\theta_{m,1}} \left(p_{ij}^{2}\right)^{\theta_{m,2}} \cdots \left(p_{ij}^{M}\right)^{\theta_{m,M}}$$

Fourth, the most important part of the demand analysis and travel forecasting problem concerns peaking; that is, the ability to differentiate travel by time of day and to measure the magnitude of peak loads, how long they last, and the extent of the accompanying congestion. No presently available methodology adequately copes with this aspect of travel forecasting, at least not when examined from a conceptual and behavioral point of view.

This is to suggest that the use of trip-purpose models, coupled with peak-hour factoring, is an unreliable technique for predicting peak-hour as well as peak-period travel conditions. Rather than attempt a critique of present-day methods and of their strengths and weaknesses, in the paragraphs that follow I shall attempt to discuss the aspects that conceptually, at least, should be incorporated in our demand models if peaking is to be reliably predicted.

Thus, it will be well to consider the various aspects contributing to or influencing the times of day at which people travel, as well as their modal choices (where they appear to be linked).

At the outset, one may hypothesize that three aspects are of prime importance to any discussion of peaking:

1. Trip purpose;
2. Institutional and physical system constraints, including transit scheduling and transport capacity; and
3. Time-of-day preference, both as related to and independent of trip purpose and institutional constraints.

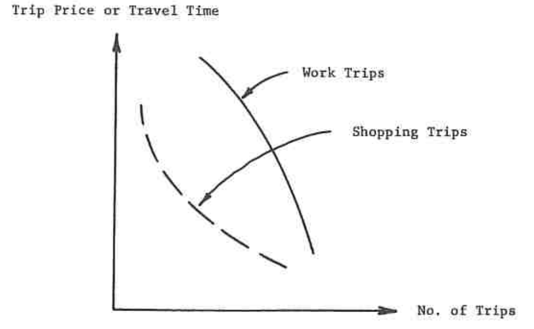


Figure 7. Hypothetical demand functions for work and shopping trips.

Stratification of demand by trip purpose clearly helps to explain trip-making behavior. Work travelers, for example, generally will tolerate more congestion, higher trip prices, and more inconvenience than will shoppers simply because the work trip will provide them with more net value (in whatever terms and whether in earnings or job satisfaction). Travelers, therefore, would suffer greater net losses by foregoing a work trip than by foregoing a shopping trip. Furthermore, one would expect work travel to be less elastic than shopping travel (i.e., the percentage change in work trips caused by 1 percent change in travel time or price should be less than that for shopping trips). Both of these hypothesized conditions are shown in Figure 7 and can be extended to all other trip purposes. A rough validation of these hypotheses can be inferred from the analysis and data incorporated in a report to the U.S. Department of Transportation (1).

Institutional and physical system constraints (here broadly defined) influence trip-making behavior in two important ways. First, work and school schedules (and any attendant flexibility) and opening-closing hours for businesses, professionals, and shops all significantly affect and limit the times of day at which trips of different purposes are made. Second, both the transit schedules and the transport system capacity can and often do constrain and influence the times at which trips are made. Changes either in the hours for various activities or in the transit schedules and available capacity can lessen or increase peaking, can either reduce or increase the total amount of trip-making, and can shift trip-making among modes.

In the same fashion, as travelers make tradeoffs among mode and route choices, based on their preferences of relative and absolute travel service and price conditions, they also must make them among different time-of-day choices. For instance, workers can often choose between getting to work on time but "fighting traffic" and getting to work early (or late) but avoiding congestion. In any case, a wide range of travel times and time-of-day scheduling choices will be available to travelers and must be matched with their preferences and tradeoffs, thus affecting both the amount and extent of peaking as well as the modal choices.

Some of the more practical situations relating to the three aspects noted previously can be explained by a number of illustrations and examples. To begin, consider the effects of increasing highway capacity. Three possibilities (or some combination thereof) come to mind:

1. As more capacity is added, the same amount of daily auto trip-making can take place with the same time-of-day distribution, thus leading to a reduction in congestion, particularly during peak hours and peak periods;
2. The same amount of daily auto trip-making can take place but some trip-makers formerly traveling before or after the peak period (of some defined length) will shift into the peak period, thus changing the time-of-day distribution; in this case, congestion during off-peak periods will be reduced and that during peak periods may or may not be reduced (depending on the extent of shifts, on volume levels, and on the capacity); and
3. An increased amount of daily auto trip-making (whether from car pool to drive-alone shifts, from "induced" trips, or from modal shifts) can take place, some or all of which can occur during the peak period; also, shifts from one time-of-day period to another can occur; congestion may or may not be reduced either during peak or off-peak periods.

TABLE 1

HYPOTHETICAL TRANSIT SCHEDULE AND TRAVEL CONDITIONS FOR ZONE i TO ZONE j TRAVELERS

Conditions	Morning Bus Schedule	Total Enroute Time (min)	Arrival at Destination		Schedule Delay[a] (min)
			Expected Time	Preferred Time	
No standees	7:30	25	7:55	8:45	+50
Standees	8:00	45	8:45	8:45	0
No standees	8:30	30	9:00	8:45	-15

[a]Schedule delay is equal to the preferred minus the expected time of arrival at one's destination.

Clearly, though, none of these possibilities (the last of which is without question the most likely one) can be forecast properly without having available models that simultaneously treat the amount of trip-making, the modal split, and the time of day at which trips are made. Furthermore, it should be evident that the first and most unrealistic of these three possibilities characterizes the assumptions implicit in our present travel forecasting techniques. The major reasons are that trip generation is not made functionally dependent on the equilibrium service, price, and performance conditions that will result from a system change, and that time-of-day preferences and constraints are not incorporated in the methodology and are not related simultaneously to trip-making and modal split.

The necessity of incorporating time-of-day preferences and constraints of the sort described earlier, and their relationship both to trip-making and modal choice, can be emphasized by more concrete examples.

Transit Example No. 1—Assume that travelers potentially going from zone i to zone j (and each having identical preferred time of arrival at zone j) are faced with the bus schedule and travel conditions given in Table 1. Even if we assume that these travelers were going to travel by bus, regardless of the conditions for travel by auto, it is hardly clear which of the three buses would be preferable and to whom—at least not without having demand functions that incorporate time-of-day and service preferences. For instance, some travelers may be willing to arrive slightly later than preferred in order to avoid the possibility of standing and the extra enroute time. Others may feel quite strongly about getting to their destination "just on time," even at the expense of spending extra time enroute and standing. Another group may be particularly impatient about enroute delays and thus choose the earliest bus even though the arrival was 50 minutes earlier than preferred. In sum, one can scarcely deal with the practical real-life forecasting problems by failing to consider the full range of service differentials or of travelers' time-of-day preferences, both of which are related to the choices available to them; nor can these problems be dealt with by simply comparing enroute travel times.

Transit Example No. 2—The bus schedules and travel conditions are identical to those outlined in Transit Example No. 1 except that the high demand for an 8:45 arrival time (relative to scheduled bus capacity) causes P percent of those trying to catch the 8:00 bus to fail in getting either a seat or standing room. Those attempting to gain space on the 8:00 bus, therefore, have to consider the probability of the bus being full, resulting in an extra 30 minutes waiting time and in their being 15 minutes late. Some travelers will be willing to gamble and accept the penalties; others will find the risk too costly to accept, thus shifting either to the earlier or later bus. Again, without knowing travelers' time-of-day preferences and without accounting for all service conditions influencing trip-making behavior, it seems unlikely that we can successfully predict how much travel will take place, when it will occur, and by what mode.

Auto Expample—The following example should demonstrate that the sorts of issues and problems that arose with the transit case (and involved the tradeoffs among time-of-day preferences, service conditions, bus schedules, and capacity) also are experienced with auto travel and thus affect both peaking and modal choices.

The auto example concerns one-way traffic flow across a bridge during the morning peak period. The following assumptions are given:

1. Fifteen hundred travelers start to work in zone A at times that are spread uniformly between 8 and 9 a.m.;
2. All workers drive alone and must cross the bridge to get to work in zone B;
3. Arrivals at the bridge entrance (or end of the queue) are uniformly spaced over time period T (in hours);
4. The bridge can service arrivals at a constant rate μ of 1,000 vph; and
5. It takes exactly 15 minutes to cross the bridge (after gaining entrance) and to complete the work trip.

Under these simplified conditions, one of the following (or some combination thereof) would result:

1. All 1,500 workers would arrive at the bridge entrance uniformly between 7:45 and 8:45. The arrival rate λ would be 1,500 vph for a time period T of 1 hour, and the service rate μ would be 1,000 vph. The average queueing delay for the arrivals during time T is approximately (2)

$$\bar{t}_q \doteq \left(\frac{\lambda}{\mu} - 1\right)\frac{T}{2} \text{ for } \frac{\lambda}{\mu} > 1$$

As a consequence, virtually all workers would suffer queueing delays, and (depending on how workers ordered themselves in the queue) all could (and most would) arrive 15 minutes late to work, on the average. In this case, the average enroute travel time of 30 minutes (to include queueing delays) would automatically incorporate the average schedule delay (or preferred minus the expected timc of arrival at destination).

2. The 1,500 workers will adjust to the potential queueing delays and thus will arrive uniformly at the bridge entrance between 7:15 and 8:45. In this case, the arrival rate will fall to 1,000 vph for an arrival time period T of 1½ hours. There will be no queueing delays, and 500 of the workers will arrive at work 15 minutes early, on the average. Note, however, that the enroute travel time of 15 minutes will not include or account for the schedule delays.

3. The 1,500 workers will adjust to the potential queueing delays and thus will arrive uniformly at the bridge entrance between 7:30 and 9:00. In this case, there also will be no queueing delays, but 250 of the workers will arrive at work about 8 minutes early and 250 of the workers will arrive 8 minutes late; note again that the enroute travel time of 15 minutes will not account for these schedule delays.

In auto situations of this sort, which are typical for peak periods in many large cities, it is not clear how the traveling public will adjust. Some will prefer to arrive either early or late to avoid congestion and queueing delays; others will decide to shift to other modes of travel or to car pooling; and so forth. But without constructing demand models that simultaneously incorporate time-of-day preferences, the full range of service conditions, and modal possibilities, it seems evident that neither modal split nor the extent of peaking can be forecast appropriately.

Inadequacy of Traditional Models—To bring these points closer to reality, one might ask why the traditional modal-split and peak factoring models are unsuitable for predicting the split and peaking, both as a general case and as applied to a city like New York. First, most if not all modal-split models and peak factoring techniques make use only of enroute travel times (including waiting times, queueing delays, and transfer times) and thus ignore the inconvenience (termed herein as "schedule delay") that results from arriving early or late to avoid or reduce congestion. Second, present models and forecasting techniques do not account for the way in which modal split and peaking are related and the extent to which they are affected by the strengths of travelers' time-of-day and service tradeoff preferences.

Turning to the first of these points, it should be evident that modal choice is not based simply on the enroute travel conditions of the alternatives and certainly does not discount the enroute travel conditions during alternative times of day. For example, why do most downtown New York workers who commute by auto travel to work during peak periods and endure extremely high enroute travel times when they could be significantly reduced if they would only travel either before or after the rush period? The answer

depends jointly on the knowledge that the commuting time could be significantly reduced only by starting to work very early or very late (relative to work starting times) and on their distaste for these other time-of-day alternatives (i.e., on a strong preference for leaving home no sooner than necessary and for arriving at work no sooner or later than necessary). On the other hand, if the peak period were not so lengthy, some peak travelers would shift to off-peak hours, the extent of switching depending of course on their time-of-day preferences and on the reduction in enroute travel time versus increase in schedule delay.

In a similar vein, one can begin to understand both modal split and peaking as well as their interrelationship. Clearly, as the level of congestion and as the length of the peak period for auto travel increases, shifts from auto to transit and from drive-alone auto to car pool will occur. Specifically, as more intense and longer peak periods occur, the amount of schedule delay generally will increase because the working, shopping, and business hours appear to remain virtually the same. Increased schedule delay, coupled with travelers' time-of-day preferences, will lead some people to shift to those facilities capable of handling higher peak loads, however uncomfortable or inconvenient they may be. In New York City, for example, it is extremely doubtful that the existing modal split (much less the level and extent of peaking) can be explained satisfactorily by making modal-split curves that employ the usual enroute travel times (to include allowances for waiting and transfers), money expenses, and income differentials. For example, transit riders in New York are probably aware that a shift to auto, in addition to entailing an arduous and lengthy trip that would permit avoidance of the subway "crush", would probably be accompanied by an early or late arrival at work.

Furthermore, it is of considerable importance to note that the data used to compare the travel times, to compute travel time ratios, and so forth, are usually incorrect. The modal-split percentages, which are based on empirical data and are incorporated in the curves for predicting future splits, often are computed on the basis of one set of data and then applied while making use of a different set. For example, empirically based modal-split percentages by trip purpose often have been calculated for travel time ratios that are based on the actual origin-destination (O-D) travel times of the travelers having that trip purpose. When this model is applied to future trip-making, different travel times are used for computing the ratios. More specifically, suppose that the empirical modal-split percentages for work trips were based on actual travel time data for work trips. The travel times then would be heavily concentrated during the peak hours, in the order of 65 to 75 percent of the total daily work trips occurring during the 4 peak hours. Given this basis for the modal-split model, it would be incorrect to use (as is often the case) off-peak or average daily travel time data or to use other travel time data for a different time period to calculate future modal splits.

Along similar lines, other inaccuracies arise because the O-D modal travel time data used in determining modal-split curves often are derived from different time periods and then are applied to still different ones. Transit work trips, for example, are usually more peaked than auto work trips. The degree of peaking depends on the extent and duration of highway congestion and on the availability of transit capacity. The O-D travel times for auto work trips thus are spread over a longer peak period than are those for transit work trips, and we can be assured that modal splits are being computed either for people having different working hours or for those having different amounts of schedule delay.

This aspect becomes of extreme importance in those situations having high and lengthy traffic congestion, particularly when considerable transit capacity is available. In Washington, D.C., for example, where transit capacity is somewhat limited (and in much the same way as auto travel is limited by congestion and street capacity), the percentage of daily transit work trips arriving at work during the morning peak hour is roughly 22 percent as compared to about 18 percent for daily auto driver work trips during the same hour (3). (Based on 1955 survey data, these results apply to work trips for the entire region rather than solely for the downtown sector; if similar percentages were available for downtown work trips, the transit percentage would probably be slightly higher and the auto percentage somewhat lower.)

In downtown New York City, however, where congestion periods are extremely intense and lengthy and which is served by very high peak-period transit capacity, about 16 percent of the daily work trips leave work during the peak 10 minutes and about 31 percent depart during the peak hour (4). (For the morning, the corresponding percentages are 10 and 31.) Although these figures represent the combined peaking for auto and transit work travel, they mainly reflect the peaking patterns for transit travel which account for almost 95 percent of the downtown work trips.

A Workable Demand Model—To formulate a workable demand model capable of incorporating the most significant of the modal-split and peaking aspects is, of course, no mean task. Moreover, the demand models, to be fully operational and meaningful, should not be formulated without considering the related problems of formulating consistent and compatible price and performance functions and of equilibrating the two sets of functions. Even so, before discussing these latter two aspects, it will be useful to propose two forms of demand models that treat peak and off-peak conditions. The first does so in a highly simplistic way; and the second, in a more satisfying and complete way (conceptually, at least).

Simplified Peak/Off-Peak Demand Model—The peak-period demand model has the form

$$q_p = \alpha_p + \beta_p t_p + \gamma_p t_o \tag{12}$$

and the off-peak-period demand model may be represented as

$$q_o = \alpha_o + \beta_o t_p + \gamma_o t_o \tag{13}$$

where q_p is the hourly volume of trips demanded during the peak period, q_o is that demanded during the off-peak period, t_p is the peak-period travel time, and t_o is the off-peak-period travel time. We would expect the parameters α_p, α_o, γ_p, and β_o to be nonnegative and β_p and γ_o to be nonpositive.

This model expresses some simple though important and logical relations. First, as congestion (i.e., travel time) during the peak period increases (while that during the off-peak period remains unchanged), some peak-period trip-making will be discouraged; some people will cancel trips altogether and others will shift to off-peak hours. If travel conditions during the peak-period are improved (but those during off-peak hours are unchanged), the peak-period flow will be increased and that during the off-peak period will be reduced. Second, both the total amount of daily flow and the split of the flow among peak and off-peak hours can change in response to changes in travel conditions during either or both of the time periods.

Composite Multimode and Time-of-Day Demand Model—Among the many ways of specifying significant influence variables, model forms, demand relations, and cross relations, the following general formulation seems sufficiently complete and logical to serve as a point of departure for further exploration and study. (In form, this demand model is not unlike the intercity and multimode passenger demand model which was developed for the Northeast Corridor by Gerald Kraft (5).)

$$q_{ij}^{m,t} = \alpha_m (Y_i)^{\beta_m} (P_i)^{\gamma_m} \left(E_{j,t}\right)^{\delta_{m,t}} \prod_{\forall x,y} \left(c_{ij}^{x,y}\right)^{\theta_{m,t,x,y}} \prod_{\forall x,y} \left(f_{ij}^{x,y}\right)^{\varphi_{m,t,x,y}} \tag{14}$$

where

$q_{ij}^{m,t}$ = quantity of trip-making between zones i and j by mode m during time period t;

Y_i = income measure for zone i residents;

P_i = population measure for zone i residents;
$E_{j,t}$ = employment measure for zone j during time period t;

$\prod_{\forall x,y}$ = the product of terms for all values of x and y ranging from 1 to M and T respectively; the expression therefore represents the product of M·T terms (of course, some elasticity values can be zero, thus reducing the number of terms);

$c_{ij}^{x,y}$ = congestion measure for travel between zones i and j by mode x during time period y;

$f_{ij}^{x,y}$ = fare or money cost measure for travel between zones i and j by mode x during time period y; and

$\beta_m, \gamma_m, \delta_{m,t}$ = the demand elasticities for mode m (or mode m and time period t) travel with respect to the income, population, and employment measures respectively. (Some of these measures will be stated in absolute terms and others in relative terms, though for this discussion it will not be necessary to be more specific.) Finally,

$\theta_{m,t,x,y}$ = elasticity of demand for mode m during time period t with respect to congestion on mode x during time period y

$$= \frac{\partial q_{ij}^{m,t} \Big/ q_{ij}^{m,t}}{\partial c_{ij}^{x,y} \Big/ c_{ij}^{x,y}}; \text{ and}$$

$\varphi_{m,t,x,y}$ = elasticity of demand for mode m during time period t with respect to fare or money cost on mode x during time period y.

The two elasticities, as x and y vary from 1 to M and T respectively, will represent the cross elasticities (i.e., they will reflect the percent change in quantity of travel demanded for one mode and time period with respect to the percent change in congestion or cost of another mode and time period). When x and y are equal to m and t respectively, however, the elasticities then will represent the direct demand elasticities. We would expect the direct elasticities to be nonpositive and the cross elasticities to be nonnegative. It is likely that empirical analysis will show that some if not many of the cross elasticity values will be zero, thus reducing materially the number of terms in the individual demand functions and the complexities to be confronted in equilibrating demand and performance functions for transport networks. For example, suppose the 24-hour day could be suitably represented by five time periods (i.e., time period 1, 7 to 9 a.m.; time period 2, 9 a.m. to 3 p.m.; time period 3, 3 to 6 p.m.; time period 4, 6 to 9 p.m.; and time period 5, 9 p.m. to 7 a.m.). For such a breakdown, it can be argued that the demand for travel during period 1, for instance, would be particularly sensitive to travel conditions during that time period and somewhat sensitive to those during periods 2 and 5. The same demand would be practically insensitive to the travel conditions during time periods 3 and 4. The cross elasticities for demand during time period 1 with respect to travel conditions during time periods 3 and 4, therefore, will be zero (or at least will be small enough to be ignored).

Equation 14 gives the demand for only one mode and time period combination out of M·T possible combinations. Thus, M·T demand functions will be required to fully specify the demand for each zonal or ij pair. Clearly, then, in situations where many modes are available the number of combinations and demand functions necessary to explain trip-making can become cumbersome, particularly when many time periods are required to reasonably explain people's time-of-day tradeoffs and preferences. Without a considerable amount of data analysis and parameter estimation it is difficult to even guess how finely the modes, submodes, and time-of-day periods should be stratified. For cities having these alternatives available, at least six modes probably should be specified: bus transit, rail transit, commuter railroad, taxi, drive-alone auto, and car pool. It seems that there are significant service and/or price differentials among these choices (that is, significant from the standpoint of influencing the amount of trip-making, the time-of-day in which trips are made, or the modal choice), and that by aggregating modes in the usual fashion (i.e., all auto versus all transit) the differentials

are made much less sharp, thus obfuscating the modal-choice question and the ability to differentiate between modes and to predict future choices.

Probably the worst aspect of this type of aggregation involves lumping drive-alone auto and car pool trips together in a single auto mode. Drive-alone auto travel has service and price characteristics that are distinctly different from both transit modes and car pool travel. Car pool travel, however, is not unlike transit travel with respect to service and price. Also, although drive-alone has service features that are all clearly superior to those for transit, car pool has some important service features that are far worse than those for transit. For instance, car poolers are restricted to a single work-trip time schedule and to a single O-D pair, whereas transit riders can take earlier or later buses to and from work and can stop off at intermediate zones or change their final destination. Aggregating the two auto submodes thus produces an average auto trip that is difficult to differentiate from a transit trip.

The problem of specifying the different modal possibilities and of defining how finely they must be stratified for forecasting purposes does not end here. Different people choose to use different modes for different parts of their door-to-door trip. For example, when comparing auto to rail transit, the service and price differentials would depend in part on whether the people traveled to the rail transit station by foot, by feeder bus, by "kiss-and-ride" auto, or by "park-and-ride" auto. Modes should be defined, therefore, by the overall modal combination for the door-to-door trip. Extending the modal-choice definition in this fashion, however, can easily triple and perhaps quadruple the number of transit modes and double the number of demand functions required. Consideration of these practical sorts of problems is hardly trivial, and careful data analysis will be required to reach firm conclusions about which modal combinations are significant and worth inclusion.

Specifying the number of time-of-day periods necessary to accurately portray the time-of-day volume variation pattern is no less difficult and certainly no less important than delineating the modal breakdown. If too few time periods are specified, it is likely that substantial inaccuracies will occur in predicting travel volumes and travel times during different times of day. This will result because the aggregated data used for estimating parameter values will mask and shift the peaks. On the other hand, if many time-of-day periods are specified, the number of demand functions will be multiplied enormously, making the task of equilibrating demand and performance functions for multimodal transport networks virtually impossible (from a computational standpoint); furthermore, the data requirements for parameter estimation would be enormous if

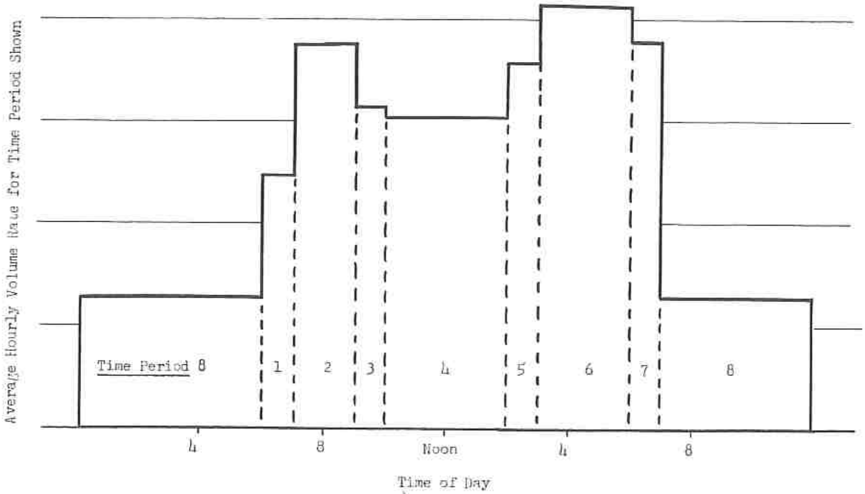


Figure 8. A suggested breakdown for time-of-day demand periods.

not out of reach for the data that presently are available. Ideally, at least eight time-of-day periods would be incorporated as shown in Figure 8. A time period on each side of the morning and afternoon major rush periods would be necessary to identify and account for shifting peak situations in which some workers (or perhaps rush-period shoppers) would go to work either early or late because of congestion during the more preferable time-of-day period. Analysis of empirical data will be necessary, however, to establish the necessity of different numbers and lengths of time-of-day periods.

Finally, it should be pointed out that to properly identify and measure time-of-day elasticities and cross elasticities, it probably will be necessary to stratify demand by trip purpose, as well as by mode and by time-of-day period. Although more clarity and accuracy will be provided by this additional stratification, the tasks of parameter estimation and of network equilibration also will be compounded considerably. No firm statements can be made about the fineness to which trip purposes should be defined; but at a minimum, two purposes—work trips and nonwork trips—should be used.

DEVELOPMENT OF PRICE-SERVICE-PERFORMANCE FUNCTIONS

The essence of the problem is to develop functions that will express the price, service, and performance conditions as a function of the facility design, vehicle technology, volume and character of usage (percent of tructs, etc.), operating and pricing policies, and so forth. Our concern is with the representation of performance and prices as viewed by the traveler.

To approach this problem, the performance function may be characterized in one of two ways:

1. Use a vector of service and price characteristics (such as time enroute, waiting time, schedule delay, and out-of-pocket money payments), or
2. Use a single price or performance variable that represents the cumulative value of money and nonmoney service and price components. In this case, it is necessary to establish commensurate values for the various components of service and price.

Whichever type of performance function is adopted, single variable or multivariate, it must be consistent and compatible with the demand function. That is, if $q = f(p,$ socioeconomic conditions, etc.), where p represents the combined money and nonmoney time, effort, and expense of travel, only a single performance function is needed or, for example, $p = f(q, C_X, \ldots)$, where C_X is a capacity measure for facility type x. However, if $q = f(p, t_1, t_2, \ldots,$ socioeconomic conditions, etc.), where p represents only the money expenses, t_1 is the access travel time, and t_2 is the line-haul travel time, then the following set of performance functions is required:

$$p = f(C_X, q, \ldots)$$

$$t_1 = f(C_X, q, \ldots)$$

$$t_2 = f(C_X, q, \ldots)$$

$$\vdots$$

Characteristics of Performance Functions

The essential aspects and characteristics of performance functions can be illustrated by using the simpler single-variable performance function rather than the more complete function involving both service and price variables. For this discussion, then, assume that

$$p = f\left(q, C_X, P_Y\right) \tag{15}$$

where C_X is a capacity measure for facility x, and P_Y represents the yth pricing policy in use for the facility. For this single-variable model, the performance or price p

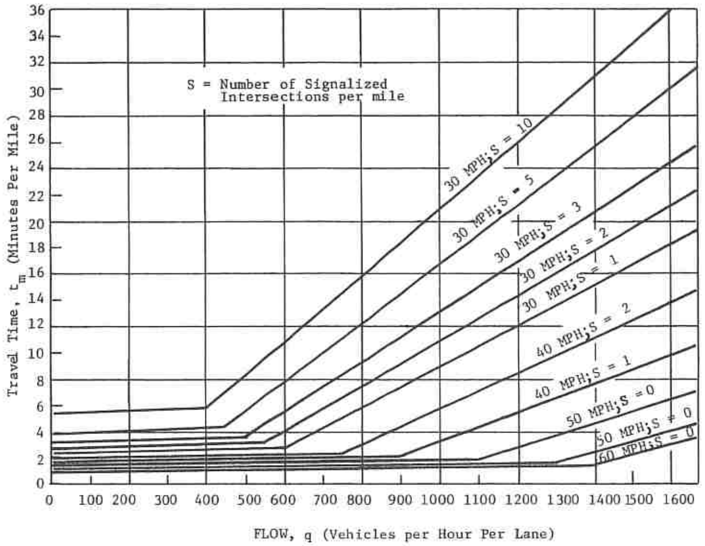


Figure 9. Capacity-restraint function curves. (From N. A. Irwin, Norman Dodd, and H. G. Von Cube. Capacity Restraint in Assignment Programs. HRB Bull. 297, 1961, pp. 109-127.)

represents the combined value (in dollars or other commensurate terms) of the money payments, time, effort, and hazards of travel, but only to the extent perceived by travelers. All travelers, however, are assumed to be homogeneous with respect to the values of the service and performance components. (This assumption is adopted merely for the purposes of illustration rather than for its realism.)

A rather typical example of a set of travel time versus volume (or so-called capacity restraint) curves is shown in Figure 9; these would be applicable when travel time is the only significant price or service variable affecting demand. The representation in Figure 9 is incomplete in two respects:

1. It fails to apply certain capacity-reducing or bottleneck types of facilities; and
2. It fails to relate travel time delay to the time interval or period over which the volume rate is sustained.

Though somewhat tangential, these two points are important enough to be clarified.

In capacity-reducing type of facilities, the service rate (or capacity) can be reduced by the shock waves produced when the traffic volume reaches a critical level. For such situations—as occurs at uncontrolled intersections and expressways, or at uncontrolled merging points—flow and the resultant performance is unstable where demand is high. Figure 10 depicts performance functions for both capacity-reducing and non-capacity-reducing types of facilities. At present, though, capacity-restraint functions used in traffic assignment have failed to represent the more complex and dynamic performance function for the former type of facility.

The second aspect—the time interval or period over which volume rate is sustained—arises partially because of a failure to differentiate between steady-state and non-steady-state queueing situations. (The effects of queues existing at the start of time periods are discussed in a later section.) To deal with the latter, the following functional form is necessary:

$$p = f\left(q,\ C_x,\ P_y,\ T\right) \tag{16}$$

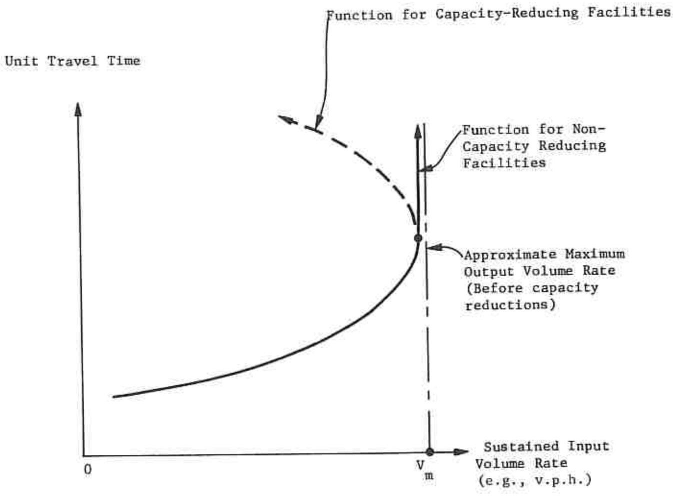


Figure 10. Approximate (non-time-dependent) relationships for travel time (or cost) versus flow for different types of facilities.

where q and C_X are stated in flow rates (e.g., vph) and T is the time interval or period over which flow rate q is sustained. In contrast, steady-state queueing, and most capacity-restraint functions, assume that λ (the arrival rate) is constant for all time and that $\lambda < \mu$ (where μ is the service rate). (The switch in notation from q to λ and from C_X to μ was made so that the terms would correspond to those common to the queueing theory literature. To avoid confusion, however, it should be emphasized that there is a one-to-one correspondence between q and λ and between C_X and μ; a differentiation and distinction is made simply because the transportation planner is more familiar with one set of definitions and notations and the queueing theorist with another.)

Figures 10 and 11 show the usual steady-state queuing model. Clearly, though, λ often does exceed μ for 1- to 2-hour rush periods (sometimes more, sometimes less) in downtown areas and on radials, thus building long queues which are worked off during later time periods when the arrival rate declines. Delays for these peak periods are not infinitely large, as implied by steady-state queueing relationships.

Thus, we need transient queueing functions, particularly ones for dealing with the exploding queue case (i.e., with the $\lambda > \mu$ case). Take a simple example in which

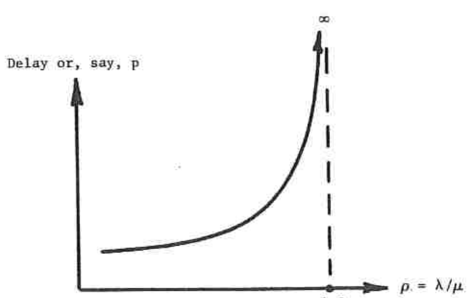


Figure 11. Generalized queueing curve for steady-state conditions.

1. Intersection capacity or service rate = μ = 1,000 vph,
2. Arrival rate = λ = 2,000 vph, and
3. Service is uniform or constant and arrivals are equally spaced.

Case 1: Let the arrival rate of 2,000 vph be sustained for a time period of ½ hour, and let there be no queue at the beginning of the time period. Then, it will take 1 full hour to clear all those arriving during the ½-hour period, and their average wait will be ¼ hour to clear the facility.

Case 2: Let the arrival rate of 2,000 vph be sustained for a time period of 1 full hour, and let there be no queue at the beginning. Then, it will take 2 full hours to clear all

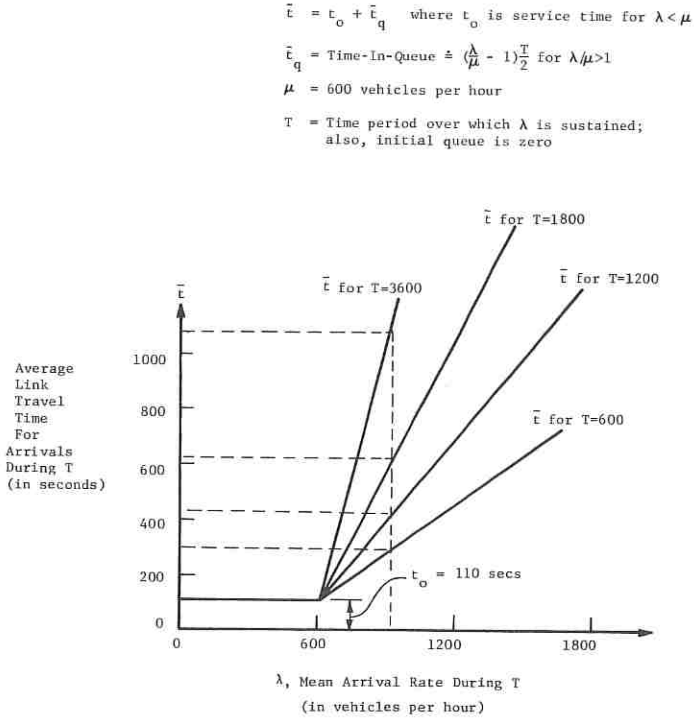


Figure 12. Time-dependent travel-time-versus-volume relationships for constant service and uniform arrival case.

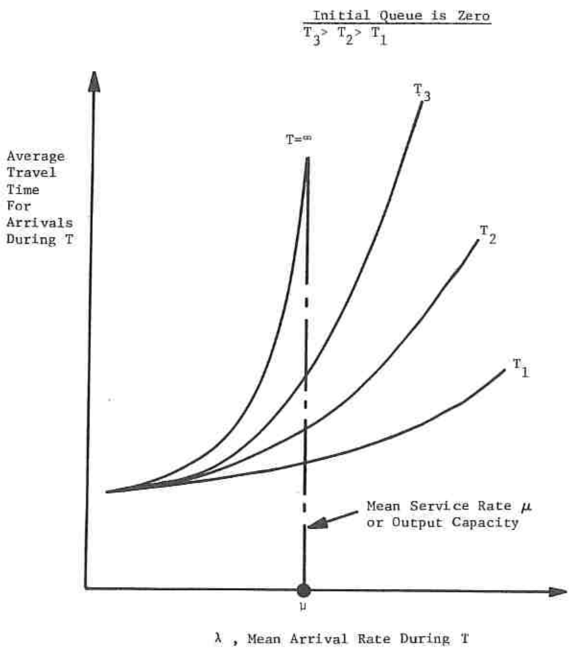


Figure 13. Time-dependent travel-time-versus-volume relationships for random arrivals and service.

those arriving during the 1-hour period, and their average wait till clearance will be ½ hour.

In short, even though λ was greater than μ (i.e., $\rho > 1$), delay was not infinite; and, quite importantly, delay obviously is a function of the time period over which the arrival rate is sustained.

Figure 12 shows these time-dependent, non-steady-state queueing relationships for the uniform service and uniform or equally spaced arrival case, and Figure 13 shows them for the random arrival and service case.

This discussion and the examples emphasize that travel time and price functions (for each link) must incorporate the time dependency and must match and be compatible with the length or time period of the corresponding demand interval. For example, if the demand function for the ith time-of-day period covers 2 hours, thenthetravel time versus

delay or capacity-restraint function must represent the travel conditions that occur for volume rates sustained over 2 hours. Of equal importance, it should be recognized that the traffic engineer, when gathering field data to establish capacity-restraint functions, must not gather data for different time periods and then factor the data to hourly rates and incorporate them into a single travel-time-versus-hourly-volume rate function. Note also that the arrival volumes and travel times of interest here are those for vehicles arriving at the upstream side of the intersection or facility during the time period in question, regardless of whether they clear the intersection or facility during that or a later time period. Also, the engineer when recording the field data should include information on the length of the queue in existence at the start of the time period.

Pricing Policy as a Determinant of the "Price" or Performance Function

It was suggested earlier that the price perceived by users was made of certain money and nonmoney payments which reflected their money, hazard, time, and discomfort "expenses." In a rough sense, one might assume that the price function now in existence on public roads and streets is equivalent to the short-run average variable cost function. Essentially, this implies the following:

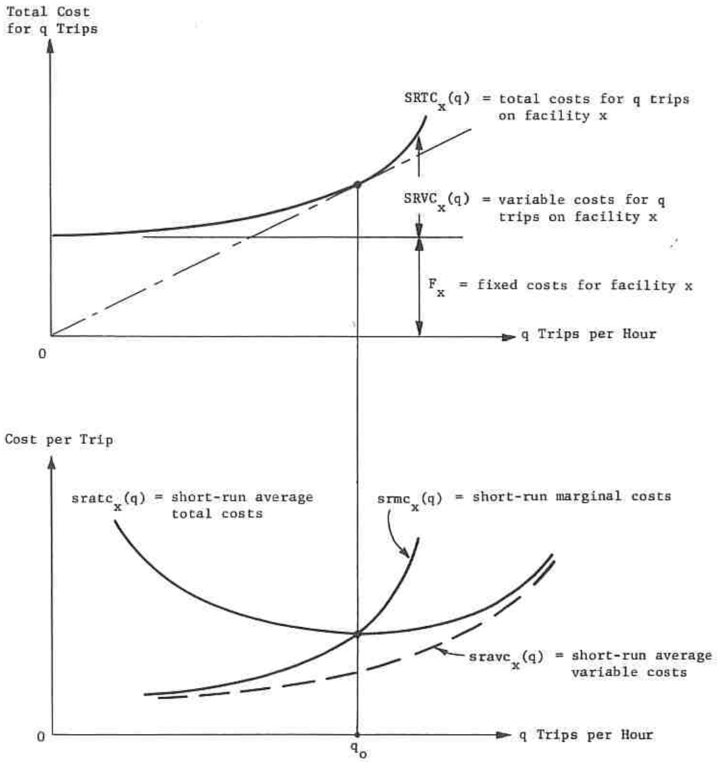
1. Perceived vehicle operating pavements ≡ variable vehicle costs;
2. Perceived parking fee payments ≡ variable parking costs;
3. Perceived user gas tax payments ≡ variable highway costs; and
4. Perceived time, effort, hazard and discomfort expenditures or payment ≡ variable time, effort, hazard, and discomfort costs.

For this discussion, short-run cost functions are those applying to time periods that are too short to alter the capacity of transport systems or links; thus the facility is fixed and neither the capacity nor the capital investment (as well as the overall travel costs for a given volume q) can be altered. In the long run, however, the facility capacity and cost relationships can be altered (both upward and downward), and thus can change the short-run cost functions. The distinction between fixed and variable costs also is important. For a particular facility, fixed costs are those that do not vary with changes in usage (i.e., with changes in q) over the short run, whereas variable costs are those that vary or change with changes in usage. Alternatively, the fixed costs may be viewed as those costs that are nonseparable with respect to nonzero volume levels and thus are common to all units of the volume using the facility.

For travel on public highways, for the existing user gas tax type of highway pricing, and for these four assumptions, the short-run average variable cost function (the $sravc_x(q)$ curve shown in Figure 14) would represent appropriately the price or performance function to be used in equilibration. Using this function presumes that travelers either do not pay or do not perceive any portion of the fixed vehicle or highway costs. Also, the curves shown in Figure 14 only apply to controlled highways and do not represent the relations applicable to uncontrolled roads and streets on which shock-wave action and backward-bending-delay-versus-volume situations often occur.

Similarly, one might argue that a comparable situation presently exists for most transit facilities and that the short-run average variable cost function can serve as an appropriate price or performance function for transit trip-making. To accept this assumption would imply that the transit fare just covers and equals the variable costs for operating and maintaining transit vehicles, trackage, stations, and maintenance facilities and that the transit system's fixed costs are covered through other revenue sources.

For other kinds of pricing policies, cost functions different from the short-run average variable cost curve would have to be used. For example, to represent the overall money and nonmoney price for highway toll facilities or for transit systems in which the toll or fare covers both the fixed and variable costs and in which the toll or fare is uniform throughout the day (that is, the fixed costs are distributed evenly among all daily users), use of the short-run average total cost function, as described by the $sratc_x(q)$ curve in Figure 14, might be appropriate. Should a peak-load pricing policy based on marginal costs be employed, then the marginal cost curve as shown by $srmc_x(q)$ would apply.



$$\mathrm{sratc}_x(q) = \mathrm{SRTC}_x(q) \div q$$

$$\mathrm{sravc}_x(q) = \mathrm{SRVC}_x(q) \div q$$

$$\mathrm{srmc}_x(q) = \frac{\partial \mathrm{SRTC}_x(q)}{\partial q} = \frac{\partial \mathrm{SRVC}_x(q)}{\partial q}$$

Figure 14. Basic short-run cost functions and relationships for a given transport facility.

The way in which these different cost and price functions can vary with facility size is shown in Figure 15.

EQUILIBRATION OF DEMAND AND PERFORMANCE FUNCTIONS

Once the demand and performance functions have been formulated, and their parameter values estimated, various techniques can be used to equilibrate them for given transport networks and pricing policies. In notational form, the task is to find the equilibrium volumes and prices (i.e., and p_i and q_i values for all n time-of-day periods) that will satisfy the following set of demand and performance or price functions:

$$q_i = f(p_1, \ldots, p_n; \text{ socioeconomic variables}) \tag{17}$$

and, for i = 1, . . ., n,

$$p_i = f\left(q_i, C_x, P_y, T_i\right) \tag{18}$$

The effect of queues in existence at the start of time periods on performance or price is discussed in the next section and also by Kraft and Wohl (6).

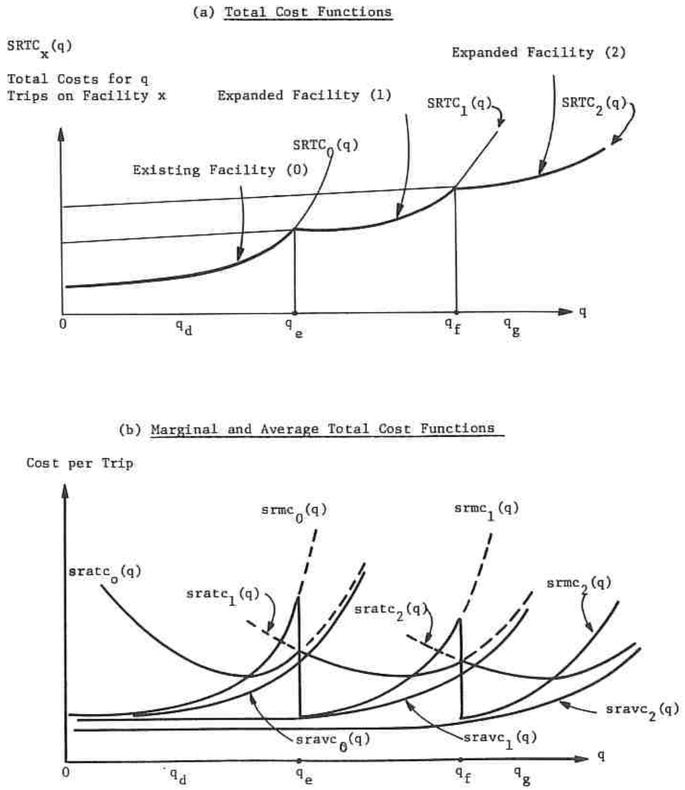


Figure 15. Short-run cost functions for three facility sizes.

It will be helpful to illustrate the interaction between these functions under different demand conditions and pricing policies; to do so, some oversimple demand and performance functions and a simple one-link unimodal transport network will be employed.

Example 1: Average Variable Cost Pricing Policy and Linear Demand

Let us consider two cases: one in which we assume that there are no hour-to-hour demand cross elasticities and that demand is linear (this last assumption is made to simplify the example); thus, for i = 1, . . . , n time-of-day periods,

$$q_i = \beta_i - \alpha_i p_i \tag{19}$$

and the second in which we assume there are hour-to-hour demand cross elasticities; thus, for i = 1, . . . , n,

$$q_i = \beta_i - \alpha_{i,\,1} p_1 - \cdots - \alpha_{i,\,n} p_n \tag{20}$$

For both equations, q_i is the quantity of trips demanded during the ith time-of-day period but expressed in trips per hour (averaged over the time period), and p_i is the average price experienced during the ith time period.

For both demand cases, a time-dependent price or performance function will be used to represent the short-run average variable cost relationship. For simplicity, let us employ the following relationship. For $q_i > C_x$,

$$p_i = p_{min} + \gamma_i\left(\frac{q_i}{C_x} - 1\right)\frac{T_i}{2} \tag{21}$$

and, for $q_i \leq C_x$,

$$p_i = p_{min} \tag{22}$$

in which p_{min} is the (constant) minimum overall trip price for unsaturated flow conditions, γ_i is the unit value of travel time and congestion during the ith time period, C_x is a capacity measure for the facility (expressed as an hourly rate), and T_i is the number of hours in the ith time-of-day period.

This clearly is an oversimple model, first because the effects of stochasticity are not incorporated, second because it suggests that the unit value of travel time and congestion is constant for all levels of travel time and congestion (i.e., an extra minute is worth the same to trips of 5 minutes as it is to those of 50 minutes), and third because it applies only to non-backward-bending flow situations. The concern at this stage of the discussion, however, is with the equilibration process and the interactions between demand and performance functions rather than with the validity of the particular models and functions employed.

Case 1: Without Hour-to-Hour Demand Cross Elasticities—For this case, congestion during the ith time-of-day period does not affect the extent of trip-making, the congestion, or the price during any other period; thus, there are no shifts of travel from one period to another. Direct solution therefore is possible, either analytically or graphically. For the former, after inverting the demand functions and equating to p_i (one unknown in each equation), the following equations (each having one unknown) are derived, and the price equation having the highest value and thus providing the smallest value of q is selected:

$$\frac{\beta_1}{\alpha_1} - \frac{q_1}{\alpha_1} = p_{min} \text{ or}$$
$$= p_{min} + \gamma_1\left(\frac{q_1}{C_x} - 1\right)\frac{T_1}{2}$$
$$\vdots$$
$$\frac{\beta_i}{\alpha_i} - \frac{q_i}{\alpha_i} = p_{min} \text{ or}$$
$$= p_{min} + \gamma_i\left(\frac{q_i}{C_x} - 1\right)\frac{T_i}{2}$$
$$\vdots$$
$$\frac{\beta_n}{\alpha_n} - \frac{q_n}{\alpha_n} = p_{min} \text{ or}$$
$$= p_{min} + \gamma_n\left(\frac{q_n}{C_x} - 1\right)\frac{T_n}{2}$$

The values of $q_1, \ldots, q_n$ can be determined directly; and those of $p_1, \ldots, p_n$, by substitution into Equation 21 where p_{min} does not apply.

Case 2: With Hour-to-Hour Demand Cross Elasticities—For this case, congestion and thus price during the ith demand period will affect flow during other periods or times of day; i.e., people are shifting from hour to hour, depending on the alternatives and their preferences, and thus the hour-to-hour and total amount of daily trip-making is affected. Stated functionally, though in linear form, the set of demand and price functions might be somewhat as follows:

1. n demand equations:

$$q_1 = \beta_1 - \alpha_{1,1}p_1 - \ldots - \alpha_{1,n}p_n$$

$$\vdots$$

$$q_n = \beta_n - \alpha_{n,1}p_1 - \ldots - \alpha_{n,n}p_n$$

2. n price equations (select the largest of each pair):

$$p_1 = p_{min} \text{ or}$$

$$= p_{min} + \gamma_1\left(\frac{q_1}{C_x} - 1\right)\frac{T_1}{2}$$

$$\vdots$$

$$p_n = p_{min} \text{ or}$$

$$= p_{min} + \gamma_n\left(\frac{q_n}{C_x} - 1\right)\frac{T_n}{2}$$

Alternatively, the n price equations can be stated as a series of inequalities as follows:

$$p_1 \geq p_{min} \text{ and } p_1 \geq p_{min} + \gamma_1\left(\frac{q_1}{C_x} - 1\right)\frac{T_1}{2}$$

$$\vdots$$

$$p_n \geq p_{min} \text{ and } p_n \geq p_{min} + \gamma_n\left(\frac{q_n}{C_x} - 1\right)\frac{T_n}{2}$$

The n demand functions and n price functions cannot be solved directly for the equilibrium q_i and p_i values because of the interdependencies that stem from the time-period-to-time-period cross elasticities. Thus, iterative numerical or programming techniques must be employed for their solution. For this simplified example, linear programming can serve as one practical technique for solving the equations simultaneously. A suitable linear programming format for accomplishing this is the following.

Determine the $q_1, \ldots, q_n$ and $p_1, \ldots, p_n$ values that will

$$\text{Maximize } Z = \sum_{i=1}^{n} k_i q_i$$

as subject to nonnegativity restrictions (i.e., all q_i and p_i values must be nonnegative) and to the following constraints:

$$q_1 = \beta_1 - \alpha_{1,1}p_1 - \ldots - \alpha_{1,n}p_n$$

$$\vdots$$

$$q_n = \beta_n - \alpha_{n,1}p_1 - \ldots - \alpha_{n,n}p_n$$

and

$$p_1 \geq p_{min}$$

$$p_1 \geq p_{min} + \gamma_1\left(\frac{q_1}{C_x} - 1\right)\frac{T_1}{2}$$

$$\vdots$$

$$p_n \geq p_{min}$$

$$p_n \geq p_{min} + \gamma_n\left(\frac{q_n}{C_x} - 1\right)\frac{T_n}{2}$$

The objective function (maximize Z) has no particular significance other than to bring about an appropriate equilibration between the demand and price functions. Also, k_i is the number of hours in the ith time-of-day period.

Example 2: Average Variable Cost Plus Uniform Fixed Facility Cost Toll Pricing Policy and Linear Demand

The pricing policy to be considered for this example is roughly equivalent to that experienced by travelers using toll facilities and, in some cases, transit systems (and may be viewed as an average total cost type of pricing policy). For this pricing policy, the overall money and nonmoney price (which reflects the total time, effort, and money expenses perceived by and expended by users) will include both the variable money and nonmoney components described in the discussion of the average variable cost pricing policy and the uniform money toll to cover fixed facility costs. Thus, p_i, the price during the ith demand period, will be equal to the toll plus the short-run average variable cost and may be regarded as roughly equivalent to the shortrun average total costs as shown earlier in Figure 14. If, for simplicity, we again assume that the time-dependent functions described by Equations 21 and 22 are suitable for representing the short-run average variable costs, then for this average total cost type of pricing policy the appropriate price or performance function for the ith time period would be as follows for $q_i \leq C_x$:

$$p_i = \text{toll} + p_{min} \tag{23}$$

and, for $q_i > C_x$:

$$p_i = \text{toll} + p_{min} + \gamma_i\left(\frac{q_i}{C_x} - 1\right)\frac{T_i}{2} \tag{24}$$

For situations in which the toll (or fare) was adjusted to raise revenues sufficiently to just cover the fixed facility costs (including interest), and in which the toll (or fare) remained uniform throughout the day, an average total cost pricing policy would result and the toll t would be

$$t = AFC_x \div 365 \sum_{i=1}^{n} k_i q_i = F_x \div \sum_{i=1}^{n} k_i q_i \tag{25}$$

where AFC_x is the annual fixed costs and F_x is the daily fixed costs for facility x, q_i is the hourly trip volume during the ith time-of-day period, k_i is the number of hours during the ith time period, and n is the number of time periods per day.

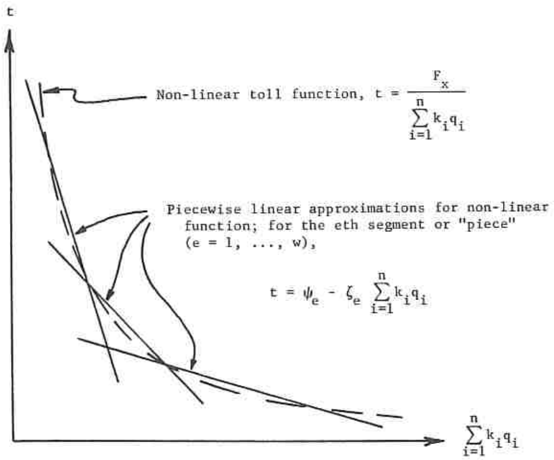


Figure 16. Linearization of nonlinear toll function.

Accompanying these price and toll functions will be demand functions that incorporate the time-of-day cross elasticities. Thus, the demand functions for the n time-of-day periods are

$$q_1 = \beta_1 - \alpha_{1,1} p_1 - \dots - \alpha_{1,n} p_n$$
$$\vdots$$
$$q_n = \beta_n - \alpha_{n,1} p_1 - \dots - \alpha_{n,n} p_n \tag{26}$$

To determine the equilibrium flows and prices ($q_1, \dots, q_n$ and $p_1, \dots, p_n$ respectively) that will simultaneously satisfy both the demand and price functions (Eqs. 23, 24, and 26 respectively) will require the use of iterative numerical or programming techniques because of the interactions resulting from time-of-day demand cross elasticities and from the toll function. Again, both these interactions and a solution technique capable of recognizing can be expressed by making use of a linear programming format. To accomplish this, however, the nonlinear toll functions must first be linearized; that is, the curvilinear toll function must be replaced or approximated by a series of piecewise linear functions. Then the price and toll functions must be expressed as inequalities, and an appropriate objective function must be chosen. Linearization of the toll function is illustrated in Figure 16; although only three linear segments or pieces were used to approximate the nonlinear function in Figure 16 (i.e., w = 3), the number of segments (w) can be increased without limit and can provide whatever accuracy is desired or necessary.

The linear programming format for this example would be as follows: Determine the $q_1, \dots, q_n$, $p_1, \dots, p_n$, and t values that will

$$\text{Maximize } Z = \sum_{i=1}^{n} k_i q_i$$

as subject to nonnegativity restrictions and to the following constraints:

$$q_1 = \beta_1 - \alpha_{1,1} p_1 - \dots - \alpha_{1,n} p_n$$
$$\vdots$$
$$q_n = \beta_n - \alpha_{n,1} p_1 - \dots - \alpha_{n,n} p_n \tag{27}$$

and

$$p_1 \geq t + p_{min}$$
$$p_1 \geq t + p_{min} + \gamma_1 \left(\frac{q_1}{C_x} - 1\right) \frac{T_1}{2}$$
$$\vdots$$
$$p_n \geq t + p_{min}$$

$$p_n \geq t + p_{min} + \gamma_n\left(\frac{q_n}{C_x} - 1\right)\frac{T_n}{2} \tag{28}$$

and

$$t \geq \psi_1 - \zeta_1(k_1q_1 + \ldots + k_nq_n)$$

$$\vdots$$

$$t \geq \psi_w - \zeta_w(k_1q_1 + \ldots + k_nq_n) \tag{29}$$

For purposes of illustration, this format will be applied to a situation that is partially hypothetical and partially empirically based. The example is intended to be applicable to a six-lane, urban toll facility which is 5 miles in length; the total annual fixed costs were set at \$1.260 million (for all 5 miles), a figure that corresponds to a facility construction and right-of-way cost of roughly \$4 million a mile. The toll function associated with this fixed facility cost is shown in Figure 17 in both its nonlinear and linear form; for the piecewise linear approximation of the toll function, the form of which is described by Equation 29, six linear segments were used.

Eight time-of-day periods were used to delineate demand. The specific times of day and the time period lengths were as shown in Figure 8, and the hypothesized demand functions for the eight time periods are given in Table 2. Although all eight functions

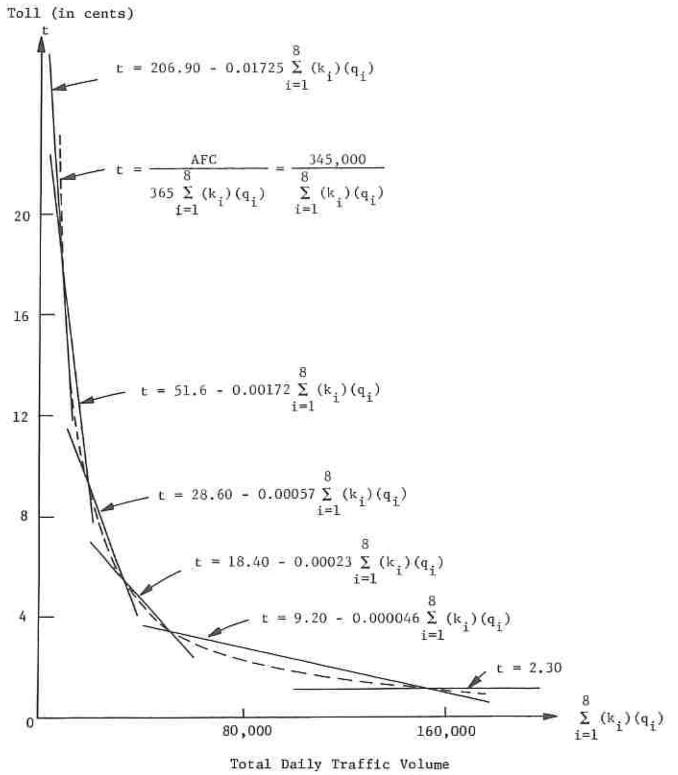


Figure 17. Linearization of average-toll-versus-daily-traffic-volume relationship for 5-mile, six-lane, urban toll facility.

TABLE 2

DEMAND FUNCTIONS FOR 5-MILE URBAN EXPRESSWAY EXAMPLE

q_1 =	$5{,}000 - 25p_1 + 19p_2$
q_2 =	$8{,}000 + 10p_1 - 21p_2 + 5p_3$
q_3 =	$7{,}200 + 6p_2 - 18p_3$
q_4 =	$7{,}500 + p_3 - 22p_4 + 2p_5$
q_5 =	$8{,}500 - 20p_5 + 4p_6$
q_6 =	$10{,}000 + 3p_5 - 25p_6 + 6p_7$
q_7 =	$8{,}600 + 10p_6 - 24p_7$
q_8 =	$3{,}300 - 10p_8$

incorporated direct demand relations (that is, the effect of the price during the ith time period on the amount of travel during the ith time period), it was not necessary to include demand cross relations (that is, to account for the effect of travel prices during other time periods on the amount of travel during the ith time period) for all cases. For example, it was assumed that nighttime travel between 7 p.m. and 6 a.m. was unaffected by peak-period travel prices and by those just before the morning peak and just after the evening peak. By contrast, travel during the morning peak-period is affected by the travel prices of both before and after the morning peak-period.

The price function used in this example represents the combined value of the toll plus the short-run average variable costs (Eq. 28); for convenience, the oversimple relationship described by Figure 12 and Equations 21 and 22 was used to represent the short-run average variable costs. (In brief, this relation ignores the effects of stochasticity; even so, neither the model structure nor the interactions are affected greatly.) The exact price functions (which embody the parameter values for γ_i and C_x) used for this purpose are given in Table 3; the values for γ_i, the unit value of travel time, congestion, and so forth, vary from time period to time period; and the value for C_x, the capacity measure, was set at 4,000 vph, and that for p_{min} was set as 64 cents.

By varying the γ_i values, it is implied that travelers during different periods react to increases in travel time and congestion to different degrees. For example, shoppers traveling during off-peak periods may be less willing or prone to tolerate some given level of congestion than will workers traveling during peak periods; or, put differently, those workers traveling to work early or late to avoid the heaviest rush periods may find travel time and increases in congestion more onerous than will those workers willing to endure the congestion occurring during the heaviest part of the rush period.

TABLE 3

PRICE FUNCTIONS FOR 5-MILE SIX-LANE (RUN 1) AND EIGHT-LANE (RUN 2) URBAN TOLL FACILITY EXAMPLE

Price Functions	For Run 1	For Run 2
$p_1 \geq 64 + t$ and	$p_1 \geq 44 + t + 0.005q_1$	$p_1 \geq 44 + t + 0.00333q_1$
$p_2 \geq 64 + t$ and	$p_2 \geq 39 + t + 0.00625q_2$	$p_2 \geq 39 + t + 0.00416q_2$
$p_3 \geq 64 + t$ and	$p_3 \geq 44 + t + 0.005q_3$	$p_3 \geq 44 + t + 0.00333q_3$
$p_4 \geq 64 + t$ and	$p_4 \geq 32 + t + 0.008q_4$	$p_4 \geq 32 + t + 0.00534q_4$
$p_5 \geq 64 + t$ and	$p_5 \geq 44 + t + 0.005q_5$	$p_5 \geq 44 + t + 0.00333q_5$
$p_6 \geq 64 + t$ and	$p_6 \geq 34 + t + 0.0075q_6$	$p_6 \geq 34 + t + 0.005q_6$
$p_7 \geq 64 + t$ and	$p_7 \geq 44 + t + 0.005q_7$	$p_7 \geq 44 + t + 0.00333q_7$
$p_8 \geq 64 + t$ and	$p_8 \geq 28 + t + 0.009q_8$	$p_8 \geq 28 + t + 0.006q_8$

Note: These functions are of the form $p_i \geq t + p_{min} + \gamma_i \left[(q_i/C_x) - 1\right] (T_i/2)$. Although the γ_i and T_i values varied from time period to time period, p_{min} and C_x were held constant and set equal to 64 cents and 4,000 vph respectively for Run 1, and to 64 cents and 6,000 vph for Run 2.

TABLE 4

EQUILIBRIUM FLOW (q_i) AND PRICE (p_i) LEVELS FOR SIX- AND EIGHT-LANE URBAN TOLL FACILITY EXAMPLE

i	Demand Period	For Run 1		For Run 2	
		q_i (vph)	p_i (cents)	q_i (vph)	p_i (cents)
1	6-7 a.m.	4,870	72.1	4,710	68.2
2	7-9 a.m.	7,270	88.2	7,470	74.3
3	9-10 a.m.	6,300	79.3	6,390	69.5
4	10 a.m.-2 p.m.	5,920	83.1	6,190	69.2
5	2-3 p.m.	7,220	83.9	7,370	72.7
6	3-6 p.m.	8,270	99.8	8,630	81.3
7	6-7 p.m.	7,550	85.5	7,650	73.6
8	7 p.m.-6 a.m.	2,620	67.8	2,620	68.2
Total daily volume		117,799		120,504	
Uniform toll		3.78 cents		4.19 cents	

Note: Run 1 applies to the six-lane case and Run 2 to the eight-lane one.

However, to illustrate the way in which the hour-to-hour and total daily trip-making can change in response to changes in facility capacity and cost, a second computer run of the linear program was made for the situation in which the facility capacity was increased from six to eight lanes and the fixed cost was increased by 14 percent. For this higher cost case, the coefficients of the linearized toll functions shown in Figure 17 were multiplied by 1.14 (i.e., increased by 14 percent), but the identical set of demand functions (given in Table 2) were used for both the six- and eight-lane cases (which will be referred to as Runs 1 and 2 respectively). Also, for the eight-lane case, the facility capacity measure (C_X) was increased from 4,000 to 6,000 vph and thus caused the price functions to change as indicated in Table 3.

The equilibrium hourly flows and prices that result from the two capacity and facility cost cases are given in Table 4. Although the shifts in hour-to-hour flows are not dramatic, some shifting and increasing peak situations do arise; and, in general, a fairly reasonable time-of-day pattern and set of interactions does appear to result. (A stronger statement in this regard obviously would depend on the use of more realistic and empirically based demand and price functions; again, the functions used herein are simplistic and included merely for illustrative purposes.)

ADAPTATION AND APPLICATION OF FORECASTING METHODOLOGY TO TRANSPORT NETWORKS

The discussion of the methodology for forecasting urban travel was kept simple; and, in addition, the discussion and models centered largely on single-link, single-mode transport networks. These simplifications were adapted primarily to simplify explication and illustration and to permit the use of straightforward graphical and analytical display and solution techniques. Even so, few fundamental changes are required to adapt these simpler concepts and models to multimodal and multilink transport systems and urban regions; mainly it is a problem of altering model notation and procedures to account for the many complexities and components of real-life systems.

To demonstrate these adaptations, the small-scale (though not necessarily oversimple) network and region illustrated in Figure 18 will be employed; the overall travel forecasting process is shown in Figure 19. (Parts a through g of the process are detailed in the Appendix).

Again, it should be emphasized that this developmental paper is concerned with forecasting travel in the short run; that is, for a situation in which the physical transport facilities and the land use, population, and employment are fixed for the point in time or year during which forecasts are to be made. This is not to say that land use and so

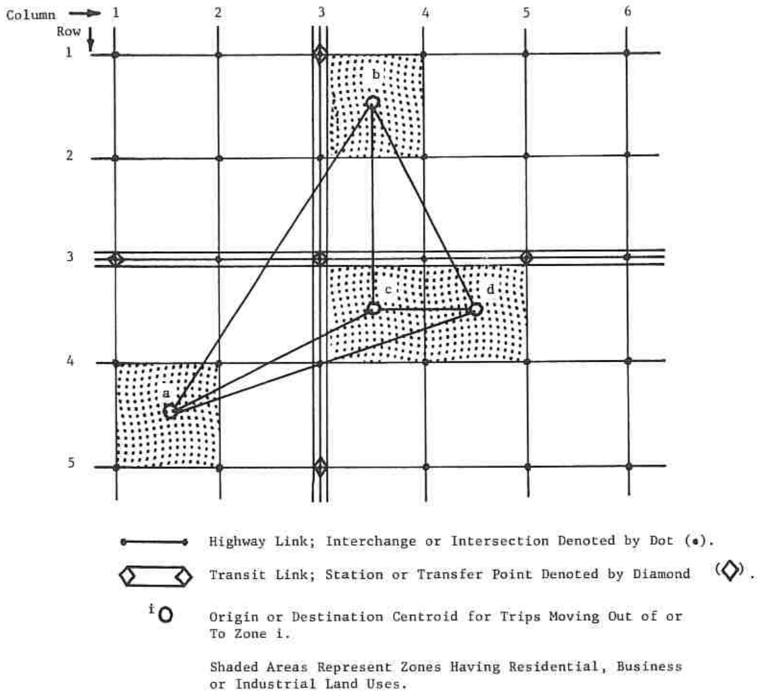


Figure 18. Hypothetical urban region and transport network at year y.

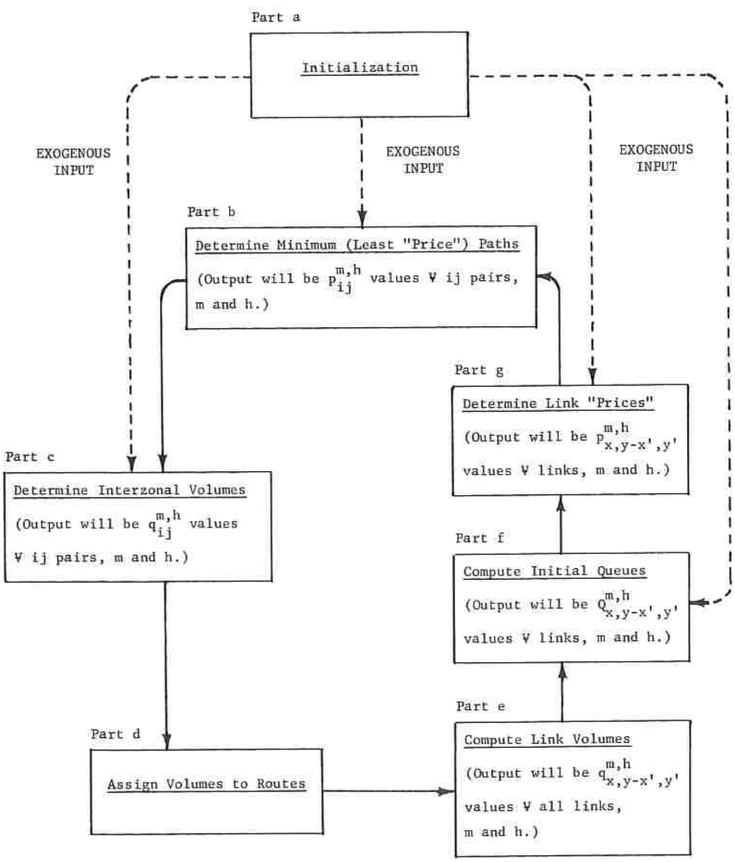


Figure 19. Travel forecasting process for multimodal transport networks.

forth are independent of transportation but to imply that the characterization of this interaction should be made within an intermediate and long-run time frame.

One objective of the travel forecasting process is to list the equilibrium flows and associated performance or price levels (or, perhaps, a vector of money and nonmoney prices if certain price and time elements are to be differentiated). If we assume that all money and nonmoney elements can be combined into a single price, the desired outputs for the system would be $q_{ij}^{m,h}$ and $p_{ij}^{m,h}$ for all values of m (the variable designating mode), for all values of h (the time-of-day period), and for all combinations of i and j (origin and destination zones) except for i equal to j.

For the system and region shown in Figure 18, there would be at least four modes—drive-alone auto, car pool, bus transit, and taxi—and thus $m = 1, \ldots, 4$. If the time-of-day demand periods outlined in Figure 8 are adopted, then $h = 1, \ldots, 8$. The number of combinations (of ij pairs) for z zones can be computed by using the following combinatorial formula:

$$\text{No. of ij pairs} = \frac{z!}{2!(z-2)!} = \frac{z(z-1)}{2}$$

However, the number of directed interzonal transfer possibilities (e.g., from i to j and from j to i) will be twice this number. For the four zones in Figure 18 having land uses, there will be six ij combinations (ab, ac, ad, bc, bd, and cd) and thus the number of directed trip pair combinations (e.g., trips from a to b) will be equal to 12. In all, there will be M times H demand functions to be used for each directed ij interzonal pair and a total of $M \cdot H \cdot z(z-1)$ demand functions to be dealt with for the region.

The total flow for a directed ij zonal pair during time-of-day period h, or q_{ij}^{h}, would be

$$q_{ij}^{h} = \sum_{m=1}^{M} q_{ij}^{m,h} \tag{30}$$

where M is the number of modes. Similarly, the total daily flow by mode m, or q_{ij}^{m}, can be determined by summing over all H time-of-day periods; that is,

$$q_{ij}^{m} = \sum_{h=1}^{H} q_{ij}^{m,h} \tag{31}$$

Neither of these flow totals, however, can be determined directly or in advance of ascertaining the equilibrium flows by mode and time of day. Nor can the total daily flow originating or ending at any zone be determined exogenously or prior to forecasting equilibrium interzonal flows by mode and time of day. Thus, q_i, the total daily trips originating at zone i, would be

$$q_i = \sum_{j=1}^{z} \sum_{m=1}^{M} \sum_{h=1}^{H} q_{ij}^{m,h} \tag{32}$$

The most difficult problems in carrying out the forecasting process (once the demand functions have been specified) are those involving the determination of the path or route that has the best price and value characteristics. There are two aspects of this phase of the forecasting process that deserve special mention. First, the prices to be used in a demand function of the form $q_{ij}^{m,h}$ = f(price for i to j trips by different modes and during different times of day; socioeconomic characteristics of residence zone i and destination zone j) are the equilibrium prices that will result from demand and supply interaction for the entire system and region. These prices also result from the

accumulated travel conditions over the various links of the travel paths between zones i and j. For example, assume that an drive-alone auto trip (m = 1) from zone a to zone b follows a route involving the links connecting intersections (2,4), (3,4), (3,3), and (3,2) as well as access between the zone centroids and the first and last intersections. (An intersection can be identified by its x and y or column and row coordinates respectively; thus, in Figure 18 intersection (4,1) is that at which the fourth column and first row intersect.) As a consequence, and if we let $p^{1,h}_{(x,y)-(x',y')}$ represent the price of traveling by mode 1 during time period h over the link between intersections (x,y) and (x',y'), the total price for trip from zone a to zone b, or $p^{1,h}_{ab}$, would be

$$p^{1,h}_{ab} = p^{1,h}_{a-(2,4)} + p^{1,h}_{(2,4)-(3,4)} + p^{1,h}_{(3,4)-(3,3)} + p^{1,h}_{(3,3)-(3,2)} + p^{1,h}_{(3,2)-b} \qquad (33)$$

The first and last terms represent the price for gaining access to or from the origin and destination zone and the initial and last links of the route.

Second, and as noted in an earlier section, if demand is not stratified by route as well as by mode and time of day, then some route assignment procedure must be adopted for assigning trips to the best path between ij pairs in an all-or-nothing fashion or for splitting trips among the different paths. The conceptually correct procedure, of course, would be to stratify demand to include route choices, but it would greatly enlarge the number of demand functions and the complexity of forecasting equilibrium flows and prices. However, by adopting some route assignment procedure as an alternative to stratifying demand by route, one assumes implicitly that neither the amount of trip-making nor the modal or time-of-day choices are affected significantly. (Short of a full-scale system analysis, the assumption will be difficult if not impossible to test.) All things considered, though, the adoption of an arbitrary route assignment procedure seems to be the wisest course; this judgment is embodied in the travel forecasting process shown in Figure 19. Although this flow chart incorporates a minimum-path route assignment procedure, such a technique was adopted merely for computational and illustrative convenience rather than as a result of any hard analysis of alternative procedures.

Several things come to mind with respect to the travel forecasting process. First, the analyst would like to know how different route assignment procedures would affect the equilibrium travel volumes and associated trip prices as well as the data processing requirements. Second, it is important to ask how much accuracy is necessary or feasible in equilibrating the demand and performance or price functions (i.e., how many iterations are necessary for satisfactory closure). Third, one must wonder whether a system of demand and performance functions will be sufficient to define a unique equilibrium (that is, whether more than one set of travel volumes and prices will satisfy the demand and performance function constraints for a given transport system and land-use pattern).

Although there are day-to-day variations in travel volumes and prices on urban transport networks (because of fluctuations in weather, people's living habits, etc.), at any one point in time and on any one day there is a unique amount of flow and level of congestion. Our problem, of course, is to predict this unique flow and to represent the variance associated with the flow and travel conditions. The latter aspect has been ignored throughout this paper, as it is a separate aspect from that of predicting the unique equilibrium values.

CONCLUSIONS

The foregoing methodology for forecasting travel on urban transport networks is offered with the sincere hope of generating both dialogue and research on the subject and of leading to analysis that will permit the development of models capable of realistically forecasting peak and off-peak travel volumes. The data gathering and processing requirements for the development of models along the lines suggested will be

formidable but necessary to develop our forecasting procedures to a higher and more fruitful level of achievement than has been possible heretofore.

It can be argued that a better or more fundamental conceptualization of the forecasting process (perhaps in the direction of that outlined herein) will be needed to improve materially our predictive capabilities. It cannot be argued with certainty, however, that the requisite effort, in the last analysis, will prove to have been worthwhile and to have improved significantly the decision-making process. For the present, and at best, one can only hope or judge that such will be the case. It is in this sense that I urge continued improvement of our forecasting methodologies and support the commitment of the research funds required for that improvement.

REFERENCES

1. Charles River Associates. An Evaluation of Free Transit Service. U.S. Department of Transportation, August 1968.
2. Wohl, M. Notes on Transient Queuing Behavior, Capacity Restraint Functions, and Their Relationship to Travel Forecasting. Papers of the Regional Science Association, Vol. 21, 1968.
3. Wilbur Smith and Associates. Mass Transportation Survey, Traffic Engineering Study, 1958.
4. Journey to Work. New York-New Jersey Transportation Agency, 1961.
5. Systems Analysis and Research Corporation. Demand for Intercity Passenger Travel in the Washington-Boston Corridor. U.S. Department of Commerce, 1963.
6. Kraft, G., and Wohl, M. New Directions for Passenger Demand Analysis and Forecasting. Transportation Research, Vol. 1, 1967, pp. 221 ff.

Appendix

TRAVEL FORECASTING PROCESS FOR MULTIMODAL TRANSPORT NETWORKS: PARTS A THROUGH G

A. Initialization

1. Land-use data for z zones,
2. Socioeconomic data for z zones,
3. Transport system definition and capacity measures for M modes and all links,
4. Pricing and/or control policies,
5. Parameter values for $M \cdot H \cdot z(z - 1)$ demand functions,
6. Parameter values for transport link performance (or price) functions, and
7. Initial estimates of equilibrium prices or $p^{m,h}_{(x,y)-(x',y')}$ values for transport system links.

B. Determine Minimum (i.e., Least Price) Paths

For each ij pair, by mode and time of day, accumulate link prices for each alternate connecting route and determine least-price path. (Output will be $p^{m,h}_{ij}$ values for all ij pairs, m and h.)

C. Determine Interzonal Volumes

Given functions of the form $q^{m,h}_{ij} = f\left(p^{1,1}_{ij}, p^{1,2}_{ij}, \ldots, p^{M,H-1}_{ij}, p^{M,H}_{ij}\right.$; socioeconomic variables$\left.\right)$, determine the volume of trips demanded between zone pairs by mode and time period for all ij pairs and for all values of m and h. (Output will be $q^{m,h}_{ij}$ values for ij pairs and for all values of m and h.)

D. Assign Volumes to Routes

Assign interzonal volumes $\left(\text{i.e., } q_{ij}^{m,h} \text{ values}\right)$, by mode and time of day, to minimum price route between ij pairs.

E. Compute Link Volumes

Determine the volumes, by mode and time of day, on the transport system links. $\Big($Output will be $q_{(x,y)-(x',y')}^{m,h}$ values for all links and all values of m and h. For modes using common facilities, such as drive-alone auto, car pool, bus transit and taxi, the modal-link volumes will be combined in commensurate vehicle or passenger flow units, whichever is appropriate.$\Big)$

F. Compute Initial Queues

For each link, by mode and time of day, compute the initial queue lengths at the start of each time period. This can be accomplished by comparing link volumes $\left(\text{or } q_{(x,y)-(x',y')}^{m,h} \text{ values}\right)$ and link capacities $\left(\text{or } C_{(x,y)-(x',y')}^{m,h} \text{ values}\right)$. $\Big($Output will be $Q_{(x,y)-(x',y')}^{m,h}$ values for all links and all value of m and h.$\Big)$

G. Performance Functions

Given functions of the form $p_{(x,y)-(x',y')}^{m,h} = f\left(q_{(x,y)-(x',y')}^{m,h}; Q_{(x,y)-(x',y')}^{m,h}; C_{(x,y)-(x',y')}^{m,h}; P_k; T_h; \text{etc.}\right)$ and given modal-capacity measures for all links $\left(\text{i.e., } C_{(x,y)-(x',y')}^{m,h} \text{ values}\right)$, pricing and/or control policies (P_k values), initial link queue lengths at start of time period $\left(Q_{(x,y)-(x',y')}^{m,h} \text{ values}\right)$, and time of day period lengths (T_h values), determine resultant link prices for all links, modes, and times of day. $\Big($Output will be $p_{(x,y)-(x',y')}^{m,h}$, values for all links and for all values of m and h.$\Big)$

The Effect of Left Turn Penalties on Traffic Assignment Accuracy

MICHAEL S. BRONZINI, Department of Civil Engineering and Pennsylvania Transportation and Traffic Safety Center, Pennsylvania State University

The objective of this study was to determine the effect of left turn penalties on the accuracy of traffic assignment results. The left turn penalty model used in the study allows the selective penalizing of left turns without penalizing right turns and through movements. The model was applied to 70 signalized intersections in the 1963, Lancaster, Pennsylvania, street network; and traffic assignment runs were made with penalty values ranging from 0 to 60 seconds. The accuracies of the assigned volumes were compared using statistical techniques. The results show that assigned traffic volumes obtained with left turn penalties in the range of 10 to 40 seconds were as accurate or more accurate than those obtained with zero penalties. Left turn penalties had more of an impact on the results of particular iterations than on average assigned volumes. The study results indicate that the application of selective left turn penalties in traffic assignment is a technique with the potential of significantly increasing traffic assignment accuracy.

•URBAN TRANSPORTATION studies have been using traffic assignment procedures for several years as tools for analyzing network traffic flows. Such procedures are based on the premise that drivers seek to minimize their travel time or cost by selecting efficient routes through a street network. In modeling this phenomenon, some version of the Moore algorithm (1) is generally employed to build minimum-path trees from an origin point to all other points in the network. Vehicular traffic is then assigned to these minimum paths.

In their original form, minimum-path-tree algorithms did not consider the extra time required for drivers to negotiate turns at intersections. More recently, several attempts have been made to correct this deficiency. The United States Bureau of Public Roads (BPR) battery of traffic assignment computer programs (2) incorporates an option for turning penalties at intersections. This method, however, is somewhat unrealistic in that no distinction is made between left and right turns, and each penalized movement in the network receives the same penalty. Two additional turn penalty models have appeared in the literature, one by Wattleworth and Shuldiner (3) and the other by Kirby (4). Neither of these models has yet been validated using data for an actual network.

A fourth turn penalty model has recently been developed at the Pennsylvania State University (PSU) (5). The PSU model was designed for selective implementation, and allows left turns to be penalized without penalizing right turns and through movements. The purpose of this study was to determine the effects of the PSU left turn penalty model on the accuracy of traffic assignment results.

THE LEFT TURN PENALTY MODEL

The turn penalty model used in this study—hereafter referred to as the PSU model—was developed with a view toward overcoming some of the theoretical and practical objections to previously proposed models. The model penalizes left turns without penalizing right turns and through movements, and allows the use of as many different left turn penalty values as there are left turns in the network.

Paper sponsored by Department of Traffic and Operations and presented at the 49th Annual Meeting.

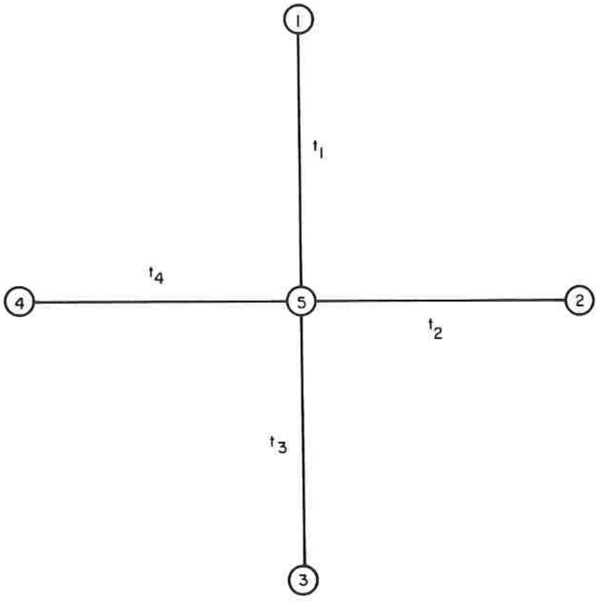


Figure 1. Standard intersection coding.

Figure 1 shows the standard method of coding an intersection. (The travel time values t_i, shown adjacent to each link, apply to both directions of travel; e.g., $t_{1-5} = t_{5-1} = t_1$.) The cumulative travel times for each of the three possible movements from node 1 are

$$Y_k = Y_1 + t_1 + t_k, \; k = 2, 3, 4 \tag{1}$$

where

Y_k = cumulative travel time from the origin node to node k, and
t_k = travel time on link k.

In general, the goal of turn penalty models is to determine new travel times Y_k', such that

$$\begin{aligned} Y_k' &= Y_k + p_{1-k}, \; k = 2, 4 \\ &= Y_k \qquad\quad\;\, , \; k = 3 \end{aligned} \tag{2}$$

where p_{1-2} and p_{1-4} are the left and right turn penalties respectively. If a turn is to be prohibited, the modeling goal is to insure that the node at the terminus of the turning movement cannot be reached via a path including both the intersection node and the node at which the movement originates. For example, if the left turn from node 1 to node 2 in Figure 1 is prohibited, the minimum-path tree should not contain the three-node sequence 1-5-2.

The intersection coding technique used in the PSU model is shown in Figure 2. The single-node representation of an intersection is discarded in favor of a system of four nodes and eight links. Right turns and through movements are accommodated on the counterclockwise intersection links, all of which are assigned travel times of zero. Left turning movements take place on the clockwise links, incurring the indicated turn penalties.

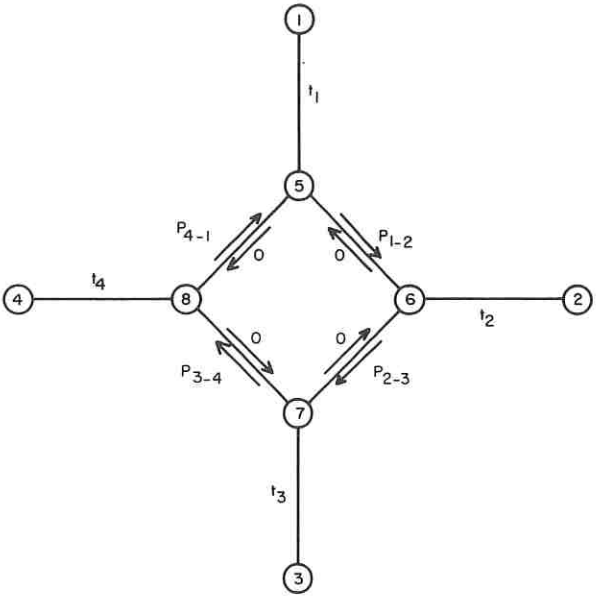


Figure 2. PSU model coding.

Thus, in extending the minimum-path tree from node 1, cumulative travel times are calculated as follows:

$$\begin{aligned} Y_2' &= Y_1 + t_1 + p_{1-2} + t_2 \\ &= Y_2 + p_{1-2} \\ Y_3' &= Y_1 + t_1 + 0 + 0 + t_3 \\ &= Y_3 \\ Y_4' &= Y_1 + t_1 + 0 + t_4 \\ &= Y_4 \end{aligned}$$

Note that these results correspond to Equation 2, if p_{1-4} is specified as zero.

To insure that the model operates as described, it is necessary to add to the Moore algorithm the stipulation that no more than two links with zero travel time may be added to the minimum-path tree in succession. Without this restriction, the path to node 2 would be determined as 1-5-8-7-6-2, with a travel time

$$Y_2' = Y_1 + t_1 + 0 + 0 + 0 + t_2 = Y_2$$

thus effectively circumventing the desired left turn penalty of p_{1-2}. With the restriction, however, link 7-6 may not be included in the minimum path from 1 to 2, because it would be the third successive link with a travel time of zero.

The PSU model provides two methods of treating prohibited left turns. Either the clockwise link corresponding to the prohibited turn may be omitted in the network description, or the turn penalty for the prohibited turn may be set at some arbitrarily high value. For example, a prohibited left turn from node 1 to node 2 in Figure 2 may be modeled by coding the link connecting nodes 5 and 6 only in the direction 6-5, or by setting p_{1-2} equal to some high value.

It is unnecessary to use the PSU model to represent all intersections in a network. The increase in network complexity occasioned by the use of the model may be limited

to those locations where actual left turn penalties are significant. Furthermore, a different penalty value may be used for each penalized left turn in the network, if desired; and penalty values may be constant or may vary with flow volume. Finally, the PSU model uses a slightly modified form of the Moore algorithm, making it readily compatible with existing traffic assignment techniques.

A complete discussion of the theory and application of turn penalties in traffic assignment and the PSU model is presented in a previous report (6).

PROCEDURE

The question of interest in this study was what effect the use of the PSU model would have upon the accuracy of assigned traffic volumes. Interzonal trips thus were assigned to an existing urban street network, utilizing a capacity-restraint assignment technique. Four different assignment runs were made, with left turn penalty values ranging from 0 to 60 seconds. For each run assigned, link volumes were statistically compared with measured average daily traffic (ADT) volumes.

Study Network

The network used for this study was the central portion of the 1963, Lancaster, Pennsylvania, street network. Computer processing time available for the study prohibited use of the entire Lancaster network. The core of the network was extracted, therefore, and traffic zones outside of this core were aggregated. This aggregation process reduced the total number of traffic zones from 271 to 178, but produced a decrease in interzonal trips of only 6.4 percent.

Network data, including an interzonal trip matrix and the coded network description, were provided by the Pennsylvania Department of Highways. A summary of network characteristics is given in Table 1.

Left Turn Penalty Values

Left turn delays at intersections were not included in the Lancaster network data. Consequently, it was necessary to generate a set of left turn penalties for this study. Two assumptions guided this generation process:

1. The PSU model should be used only at signalized intersections.
2. Turn penalty values should vary with different levels of intersection flow volume.

These assumptions were based upon previous experience with the PSU model (5).

Examination of the data for the Lancaster network revealed that there were 70 signalized intersections applicable to the PSU model. To adhere to the second assumption, it was necessary to stratify these intersections by intersection volume levels. The particular volume of interest was the approach volume conflicting with left turn maneuvers. Because the only volume information available was in terms of ADT, however, a different stratification variable had to be used. The variable selected was the average ADT for opposite intersection legs. This variable includes all through traffic at an intersection; thus, the variable is, in some sense, a measure of the volume level that a left turn movement must cross.

Calculation of average ADT's for all opposite approach pairs where left turns conflicted with other movements produced a set of 98 average ADT values. Several frequency distributions of these values were prepared, using varying class intervals. The distribution given in Table 2 was selected and used as the basis of the left turn penalty value classification scheme implied by the second assumption. The selected distribution defined three classes of approximately equal size and a relatively small, low-volume class. All left

TABLE 1

STUDY NETWORK CHARACTERISTICS

Urban area	Lancaster, Penn.
1963 population	103,900
Area	48 sq mi
Dwelling units (1963)	36,400
Interzonal trips	239,776
Traffic zones	178
Total nodes	632
Total links	1960

TABLE 2

FREQUENCY DISTRIBUTION OF AVERAGE ADT'S

Class	Average ADT Range	Number of Approach Pairs
1	0-3,000	9
2	3,001-6,500	31
3	6,501-10,000	31
4	Over 10,000	27
Total		98

TABLE 3

LEFT TURN PENALTIES

Turn Penalty Set	Left Turn Penalty Values (sec)			
	Class 1	Class 2	Class 3	Class 4
0	0	0	0	0
T	0	10	15	20
2T	0	20	30	40
3T	0	30	45	60

turn movements in the network that were to be penalized were assigned to one of these four classes.

Finally, the particular turn penalty values to be applied to each class were selected. The primary concerns in this selection process were to obtain penalty values that would span a wide range of reasonable left turn delays and to provide systematic variation of the values. Table 3 gives the selected left turn penalties. As shown, the low-volume class received a zero penalty (input as 0.001 minute) throughout, and the relative magnitudes of the penalties for classes 2 to 4 were held constant.

All classes initially were given a penalty value of zero, to provide a set of assignment results against which to assess the effects of the PSU model. That is, the results obtained with turn penalty set zero were identical to those that would be obtained using standard intersection coding. Turn penalty set T includes in its range the values of 12 and 18 seconds, which have commonly been used in conjunction with previous turn penalty models. The remaining two sets were derived by doubling and tripling the T values.

Implementation of the PSU model at the 70 selected intersections required the addition of three nodes and eight links at each location. The final coded network thus contained a total of 842 nodes and 2,520 links.

Traffic Assignment Procedure

The choice of a traffic assignment procedure was limited to one that uses the input provided for the study network. Because the traffic assignment phase of the Lancaster study was performed with the BPR battery of traffic assignment computer programs, the BPR procedure was chosen for use in the present study. This was the only logical choice, because selection of a different procedure would have required a somewhat arbitrary data conversion process.

The BPR assignment procedure features all-or-nothing loading, with as many iterations of capacity restraint as desired. A complete description of this procedure may be found in the Traffic Assignment Manual (2).

Although the BPR procedure was used in this study, the BPR computer program package was not used. Instead, the traffic assignment program developed to implement the PSU model (5) was modified to incorporate the essential features of the BPR procedure.

Traffic Assignment Computer Runs

Four traffic assignment runs were made, one for each set of left turn penalty values. A free assignment and three iterations of capacity restraint were specified; thus each run produced four sets of assigned link volumes. The only input data that changed from one run to the next were the turn penalty values. Total computer processing time amounted to 3,690 seconds per run.

Analysis of Assignment Accuracy

Several techniques for comparing assigned link volumes with measured link volumes were used. These included the computation of weighted error by volume group, root mean square (RMS) error comparisons, simple linear regression, and link volume frequency analysis.

The weighted error computation is explained fully by Humphrey (7). Essentially, it involves calculating the percent standard error in the assigned volumes for a number of link volume groups and weighting these errors by the percentage of the total link volume contained in each group. This statistic is calculated as

$$E_j = \frac{100\, s_j}{\overline{q}_j} \frac{m_j\, \overline{q}_j}{\sum_{i=1}^{N} q_i} = \frac{100\, m_j\, s_j}{\sum_{i=1}^{N} q_i} \tag{3}$$

in which

E_j = weighted error for volume group j,
$\overline{q}_j$ = average ground count volume for volume group j,
m_j = number of links in volume group j,
q_i = ground count volume for link i,
N = number of links with a recorded ground count, and
s_j = standard error in assigned volume for volume group j.

The standard error s_j is calculated as

$$s_j = \sqrt{\frac{\sum_{i=1}^{m_j} (q_c - q_a)_i^2}{m_j - 1}} \tag{4}$$

where

q_c = ground count link volume,
q_a = assigned link volume, and
i = observation number.

The summation of the weighted errors is a relative index of the accuracy of the assigned volumes. This is computed as

$$E_t = \sum_{j=i}^{n} E_j \tag{5}$$

in which

E_t = total weighted error in assigned volumes,
E_j = weighted error for volume group j, and
n = number of volume groups.

The simple linear regression technique consists of estimating the parameters of the following equation:

$$q_a = A + Bq_c \tag{6}$$

where q_a and q_c are as defined previously. This technique produces estimates of A and B; the standard deviations of these estimates; and the coefficient of determination R^2, which indicates the percentage of the variation in q_a which is "explained" by variation in q_c. Simple linear regression thus is useful for investigating the degree of correlation between the two variables.

The fourth technique used was link volume frequency analysis. This consists of counting the number of links having assigned volumes in various volume groups and comparing these totals with those obtained using ground count volumes.

The study network contained 559 links for which measured ADT volumes were available. The assigned volumes from all four iterations for each run were compared with the ADT volumes using the techniques discussed earlier. Average assigned volumes, obtained by summing the volumes assigned to a link for each iteration and dividing the total by the number of iterations, were also compared with the measured values.

RESULTS

Table 4 gives summary statistics for each iteration of each assignement run. These results show an average underassignment of 11 to 15 percent, which is typical of assigned traffic volumes.

Weighted Errors in Assigned Volumes

Figure 3 shows the effect of the left turn penalty value on the total weighted error in assigned volumes. In general, the total weighted error decreased as the left turn penalty increased to an optimum penalty value, which was in the range of T to 2T. The smallest error occurred with fourth iteration volumes and turn penalty set 2T, whereas the sharpest reduction in weighted error occurred with third iteration volumes in going from zero turn penalty to penalty set T. The largest decrease in weighted error from iteration one to iteration four occurred with turn penalty set 2T, and the smallest reduction was obtained with zero turn penalty values. Note, however, that the lowest weighted errors occurred with average assigned link volumes where the effect of the turn penalty was negligible.

Root Mean Square Errors in Assigned Volumes

In Figure 4, the overall RMS error is plotted against the turn penalty value. As might be expected, these curves are similar in shape to those of Figure 3. A 10 percent reduction in the RMS error occurred for the same two volume sets that showed

TABLE 4

SUMMARY OF ASSIGNMENT RESULTS

Iteration Number	Turn Penalty Set	Average Assigned Volume (vpd)	Average[a] Difference (vpd)	Total Weighted Error (percent)	RMS Error (vpd)
1	0	6,201	-1,021	41.7	3,126
	T	6,202	-1,020	41.5	3,117
	2T	6,278	-944	43.8	3,278
	3T	6,358	-864	43.2	3,236
2	0	6,397	-825	44.5	3,349
	T	6,166	-1,056	42.0	3,171
	2T	6,177	-1,045	41.6	3,130
	3T	6,312	-910	43.3	3,217
3	0	6,262	-960	42.0	3,201
	T	6,311	-911	37.3	2,852
	2T	6,326	-896	37.8	2,869
	3T	6,391	-831	36.8	2,842
4	0	6,191	-1,031	40.5	3,033
	T	6,246	-976	37.6	2,853
	2T	6,341	-881	35.9	2,691
	3T	6,330	-892	37.6	2,839
Avg	0	6,262	-960	27.8	2,148
	T	6,231	-991	28.4	2,221
	2T	6,280	-942	28.2	2,194
	3T	6,348	-874	28.0	2,167

[a]Average difference = average count − average assigned volume, = 7,222 − average assigned volume.

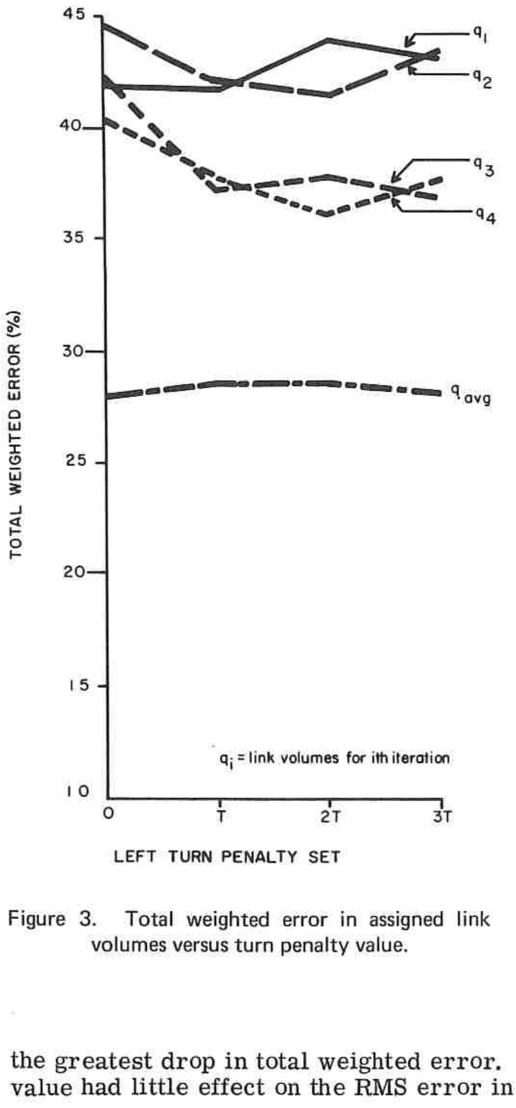


Figure 3. Total weighted error in assigned link volumes versus turn penalty value.

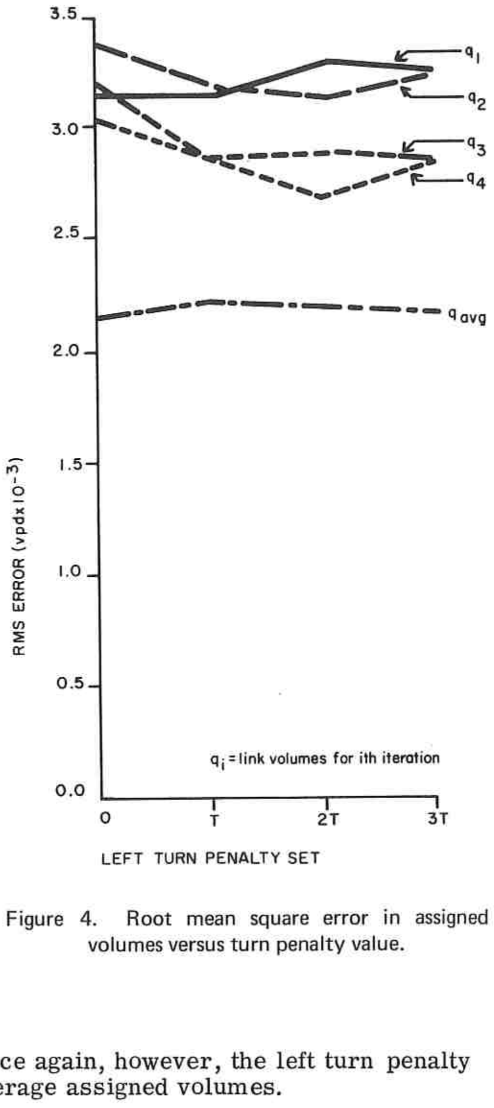


Figure 4. Root mean square error in assigned volumes versus turn penalty value.

the greatest drop in total weighted error. Once again, however, the left turn penalty value had little effect on the RMS error in average assigned volumes.

Simple Linear Regression

The results of least-squares regression analysis, in which assigned link volumes were compared with ground count volumes by estimating the parameters of Equation 6, are given in Table 5. As shown by the standard errors in the parameter estimates for a given iteration number, the left turn penalty did not have a statistically significant effect on the estimated slope and intercept values. The variance of the parameter estimates generally decreased slightly as the turn penalty increased from zero to 2T.

Given these results, the coefficient of determination R^2 is of particular interest because it shows how closely the data are grouped around the regression line or, in this case, the degree of correlation between the assigned and observed link volumes. Figure 5 shows a graphical comparison of R^2 values for differing turn penalty values. These results are very similar to those presented previously. The greatest improvements in R^2 values occurred with the same link volume sets that showed the largest reductions in weighted and RMS errors. Once again the curve for average link volumes deviates only negligibly from the horizontal.

TABLE 5

LINEAR REGRESSION RESULTS

Iteration Number	Turn Penalty Set	Intercept[a] Value (vpd)	Intercept[a] Standard Error (vpd)	Slope[a] Value	Slope[a] Standard Error	R^2
1	0	225	225	0.823	0.026	0.636
	T	538	220	0.784	0.026	0.624
	2T	631	236	0.782	0.028	0.590
	3T	577	237	0.801	0.028	0.600
2	0	242	252	0.852	0.029	0.600
	T	124	230	0.837	0.027	0.635
	2T	176	226	0.831	0.026	0.640
	3T	160	239	0.852	0.028	0.625
3	0	-260	240	0.903	0.028	0.651
	T	-361	213	0.924	0.025	0.712
	2T	-369	214	0.927	0.025	0.710
	3T	-478	215	0.951	0.025	0.720
4	0	-544	225	0.932	0.026	0.693
	T	-458	211	0.928	0.025	0.717
	2T	-621	201	0.964	0.024	0.750
	3T	-467	213	0.941	0.025	0.720
Avg	0	-77	146	0.878	0.017	0.823
	T	-39	150	0.868	0.018	0.814
	2T	-46	151	0.876	0.018	0.816
	3T	-52	152	0.886	0.018	0.817

[a]Volume assigned = A + B x ground count volume; 559 observations per equation.

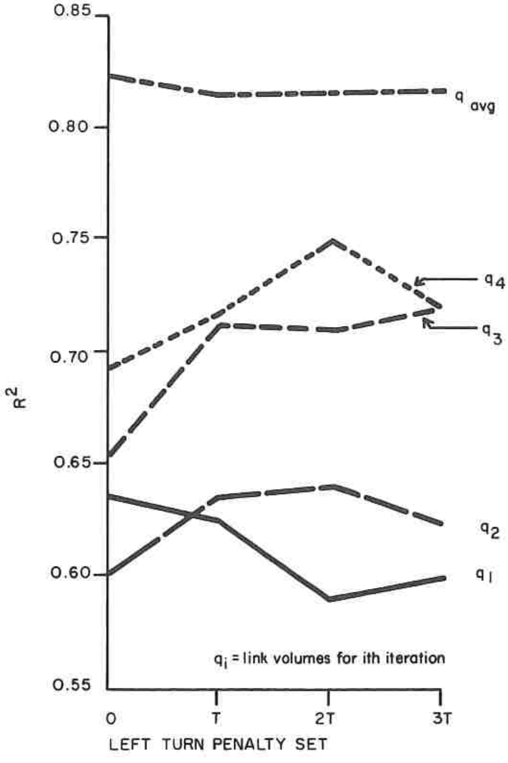


Figure 5. Coefficient of determination versus turn penalty value.

TABLE 6

LINK VOLUME FREQUENCIES

Volume Range (vpd x 10^{-3})	Count	Turn Penalty Values 0	T	2T	3T
Fourth Iteration Volumes					
0-0.5	0	44	20	24	18
0.5-1	8	41	33	28	34
1-2	55	80	91	84	71
2-3	64	41	59	64	69
3-5	95	69	67	68	102
5-10	186	148	168	167	138
10-15	127	107	79	77	71
15-20	17	24	34	39	51
20-25	6	5	8	8	5
25-30	0	0	0	0	0
Over 30	1	0	0	0	0
Total	559	559	559	559	559
Average Volumes					
0-0.5	0	1	0	0	0
0.5-1	8	18	8	11	17
1-2	55	78	84	79	70
2-3	64	87	80	82	73
3-5	95	84	100	100	110
5-10	186	175	171	165	167
10-15	127	94	91	100	92
15-20	17	17	20	20	27
20-25	6	5	5	2	3
25-30	0	0	0	0	0
Over 30	1	0	0	0	0
Total	559	559	559	559	559

Link Volume Frequencies

Table 6 gives the number of links having assigned volumes in various volume groups for fourth iteration volumes and average volumes, as well as for ground count volumes. In general, the effect of the PSU model was to produce more accurate link volume frequencies in the lower volume ranges, less accurate ones in the middle volume ranges, and frequencies in the high volume ranges that did not differ significantly from those obtained without the model. The same general pattern is true for both fourth iteration volumes and average volumes.

ANALYSIS OF RESULTS

Model Performance

The effect of the PSU model on the accuracy of assigned link flow volumes depends to a large extent on which set of assigned volumes is considered. Use of the model increased the accuracy of the assignments for third and fourth iteration volumes, which are used in many urban transportation studies as the predicted link flows. On the other hand, the model had little effect on the accuracy of average assigned volumes, which are also widely used in practice.

Perhaps the more important result to note is that the PSU model did have an effect on assigned volumes. This is true both for volumes resulting from a particular iteration and for average volumes, as demonstrated by the link volume frequencies. Note also that the use of the model, in most instances, produced either beneficial or negligible changes in the various accuracy statistics. The principal findings of this study thus may be stated as follows:

1. Minimum paths determined with the use of left turn penalties differed from those determined with zero penalties.
2. Assigned traffic volumes obtained with left turn penalties were as accurate or more accurate than those obtained with zero penalties.

Cost-Benefit Considerations

The practicing transportation planner is interested in whether or not the cost of implementing a turn penalty model is justified by the benefits it may provide. This is a difficult question to answer, primarily because of a lack of any recognized method of evaluating the benefits of a modeling technique.

The results of this study do serve to validate the basic cost-benefit concept incorporated into the structure of the PSU model. This is the concept of selective implementation. Recall that the structure of the PSU model is such that the model need not be employed in describing every intersection in a network. Indeed, in the present study, the model was used at only 70 of the 454 possible nodes. The use of the model at this limited number of locations, however, did produce changes in minimum paths, with the effects discussed previously. The results of this investigation thus indicate that the use of selective left turn penalties, and the attendant limiting of network coding complexity and model implementation costs, is a valid traffic assignment technique.

Left Turn Penalty Values

The left turn penalty values that produced the most significant improvements in assignment accuracy were generally turn penalty sets T and 2T, which encompass penalties in the range of 10 to 40 seconds. Because commonly used values are at the lower end of this range, this result indicates that better assignment accuracies might be obtained with existing turn penalty models by using somewhat higher turning penalty values. Ideally, of course, actual measured turn penalty values should be used. Difficulties in measuring appropriate values, however, may force continued reliance on surrogate techniques such as those used in this study. If such is the case, the results concerning penalty values may prove to be immediately useful.

The results also have implications concerning the procedure of this investigation. First, the turn penalty values used apparently bracketed the optimum penalty values. Hence this study constituted an adequate test of the sensitivity of the assignment results to the left turn penalty value. Second, the results of the study validate the assumption made in developing penalty values; i.e., that the magnitude of the left turn penalty should vary with intersection flow volume. This statement must be tempered somewhat, in that the opposite condition, i.e., the use of a single penalty value for each assignment run, was not tested. The results, however, do show that the assumption led to accurate predictions of link flow volumes.

Turn Penalties in Traffic Assignment

Several proposed and operational turn penalty models were reviewed prior to development of the PSU model. Because the PSU model is similar to these proposed models in many respects, the results of this study have several implications concerning the use of turn penalty models in general.

The results show that the use of turn penalties in traffic assignment is a modeling technique that has the potential of increasing the accuracy of traffic assignments. Development, refinement, and implementation of such models thus are in order and should be continued.

The results of this study also indicate that the flexible type of turn penalty model appears to be the most fruitful. That is, turn penalty models that are capable of selective implementation and that allow different penalties for right and left turns appear to be more promising. This is so because the use of models with the opposite characteristics will lead, in general, to use of turn penalty values of a relatively lower order of magnitude than may be appropriate for left turn movements, as a result of the equal penalizing of left and right turns.

CONCLUSIONS

The principal findings and conclusions of this investigation are as follows:

1. Minimum paths determined with the use of left turn penalties differed from those determined with zero penalties.
2. Assigned traffic volumes obtained with left turn penalties were as accurate or more accurate than those obtained with zero penalties.
3. The use of left turn penalties appears to have more of an impact on assigned volumes for particular iterations than on average assigned volumes.
4. The procedure of selectively penalizing left turn movements is a valid and useful traffic assignment technique.
5. Improved traffic assignment accuracies might be obtained with existing turn penalty models if penalty values were made somewhat higher than those commonly used in the past.
6. The use of turn penalty values that vary with intersection flow volume level appears to be a valid procedure.
7. The application of turn penalty models in traffic assignment is a modeling technique that has the potential of significantly increasing assignment accuracy. The development and implementation of such models thus should be continued.

ACKNOWLEDGMENTS

This study was made possible through the support of the Pennsylvania Transportation and Traffic Safety Center at the Pennsylvania State University. Computer processing time was provided by the University. The Urban Transportation Studies Division, Bureau of Advance Planning, Pennsylvania Department of Highways, provided the data for the study. Many helpful suggestions in the conduct of the research were received from Thomas B. Davinroy, Joseph L. Carroll, Thomas D. Larson, and Benjamin A. Whisler, all of the Pennsylvania State University.

REFERENCES

1. Moore, E. F. The Shortest Path Through a Maze. Internat. Symp. on the Theory of Switching, Harvard Univ., April 1957. Harvard University Computation Laboratory Annals, Vol. 30, 1959, pp. 285-292.
2. Traffic Assignment Manual. U. S. Bureau of Public Roads, Office of Planning, Urban Planning Division, U. S. Govt. Printing Office, Washington, D. C., 1964.
3. Wattleworth, J. A., and Shuldiner, P. W. Analytical Methods in Transportation: Left Turn Penalties in Traffic Assignment Models. Proc. ASCE, Jour. Engineering Mechanics Division, Vol. 89, Part 1, No. EM6, 1963, pp. 97-126.
4. Kirby, R. F. A Minimum Path Algorithm for a Road Network With Turn Penalties. Proc. Third Conf. Australian Road Research Board, Vol. 3, Part 1, 1966, pp. 434-442.
5. Davinroy, T. B., Bronzini, M. S., and Herendeen, J. H., Jr. Transportation Network Analysis Utilizing Selective Left Turn Penalties. Pennsylvania Transportation and Traffic Safety Center, Pennsylvania State Univ., University Park, July 1969.
6. Bronzini, M. S. The Effect of Left Turn Penalties on Traffic Assignment Results. Pennsylvania Transportation and Traffic Safety Center, Pennsylvania State Univ., University Park, May 1969.
7. Humphrey, T. F. A Report on the Accuracy of Traffic Assignment When Using Capacity Restraint. Highway Research Record 191, 1967, pp. 53-75.

Characteristics of Urban Activity Patterns

GEORGE T. LATHROP, Department of City and Regional Planning,
University of North Carolina at Chapel Hill

•THIS PAPER describes the continuing research on the daily activity patterns of residents of a metropolitan area, which is being conducted within the Department of City and Regional Planning at the University of North Carolina.

Despite the tremendous gains made in the past 20 years in understanding and simulating the urban area, it is apparent that techniques are tied strongly to observations of present patterns and draw very little from a theoretical understanding of the city, its processes, or the forces that lie behind the observable phenomena.

The work described here focuses on activities as the means for characterizing urban phenomena. These activities are simply the things people do—what, where, when, and how long. Their collection or sum is the pattern we observe in the urban area.

Work to date has concentrated on defining satisfactory measures of activity patterns. Although several measures have been used (described later) that strongly resemble those used in various analyses of travel, it is anticipated that measures more completely describing the complexity of daily activities can be defined. One such measure is suggested.

Implicit in this work is a model of the activity patterns of urban residents. The analysis suggests that there are both strong similarities and strong differences in the activity patterns of various groups of urban residents. These patterns are differentiated by familiar socioeconomic variables and by measures of the residents' accessibility to specialized locations for urban activities. It is anticipated that the model may be responsive to changes in these patterns.

The paper is organized into three sections. The first briefly discusses a theoretical framework for the analysis and for the model in general. The second describes more specifically some of the relationships anticipated between activity patterns and characteristics of various urban subgroups. The third section discusses the results of analysis to date, concentrating on socioeconomic groups and using crude measures of mobility and spatial effects.

FRAMEWORK

The activities in which people engage are, in effect, their choices in a marketplace that offers various opportunities for the dispersal of their resources—both monetary and otherwise.

Many of the fundamental notions about activities incorporated here are from Stuart Chapin and his colleagues (1). In a recent article he has outlined a working schema for the development of a conceptual framework for urban spatial structure using activity analysis and activity systems. Simply stated, his schema is an evolutionary process involving motivation-choice-activity.

In discussing the components of the schema, he suggested that motivation is derived from two sets of needs, fundamental and supplemental. Fundamental needs are involved with shelter, clothing, food, and so forth; i.e., choices to minimize feelings of discomfort or deprivation. Supplemental needs for reason of achievement and status are requisite to a "full sense of well-being" and require choices to maximize satisfaction. He suggests that each of these needs (and they may be further broken down)

Paper sponsored by Committee on Land Use Evaluation.

is satisfied by or is sought for in different roles and arenas, sometimes simultaneously and sometimes separately.

In discussing the choice component, Chapin specifies the activity as the output, with motivation (and the values it represents) forming the inputs to a final decision. "The context is the social system consisting of the environment and all other human activity relevant to the situation in hand" (1). He suggests that in making choices of activities and in budgeting his time, the individual attempts to find an optimal combination based on his needs for "security, achievement, status, and other needs essential to his sense of well-being." The final output is, of course, his set of activities in various time scales: daily, weekly, annually, and throughout his entire life cycle.

In examining the activity patterns that result from the motivation-choice-activity sequence, it is obvious that all three elements of the sequence must be influenced by the alternatives available. The alternatives are obviously numerous. One clear distinction may be drawn between those that occur in the home and out of the home. Chapin, through the use of time-budgets, has begun an investigation of this division and of the various activities that make up each subset (2). Activities, however, are those with which planners—transportation and otherwise—are directly concerned, for these are the activities that crowd streets and highways, that become patronage when translated into market transactions, and that create demands for publicly supplied or regulated facilities and services. In brief, the patterns observed resulting from the motivation-choice-activity sequence of urban residents are fundamental factors in many decisions that affect the physical structure of the urban area.

At the same time, the components of the physical structure—the spaces adapted to various activities—exert an influence on the choices made. The situation is strongly analogous to the economic concepts of supply and demand and their interaction with price and production.

Activities may be distinguished according to who performs them. Chapin suggests that three groups—residents, firms, and institutions—are distinguishable. This discussion involves the activities of the residents.

Finally, activities are clearly cyclical. Some are daily—the trip to work, eating, some sort of recreation. Others may be weekly—grocery shopping, trips to the bank. Still others may be monthly, annual, or at even longer periods (for example, the decision to move.) A general discussion of cyclical activity is given by Chapin (3).

Activities are of varying duration: Some have fixed durations; others, flexible. Watching television is of variable duration, but going to a movie usually involves a commitment of at least 2 hours.

Our focus is on the linked sequence of activities occurring over a day. Two kinds of linkage sequences occur. First is the set of out-of-home activities comprising the full daily cycle. Second is the smaller set of activities that occur on each foray out of the home—activities that are linked together on each individual journey.

EXPECTED RELATIONSHIPS

In attempting to structure the choice of activities for analysis (and eventually for simulation) there are apparently three principal dimensions or components of activity choice and sequencing in addition to the socioeconomic characteristics of the individual. These are time, space, and the activities themselves. The selection of daily activities, the selection of places where they will be performed, and the time devoted to each activity are all interconnected; together they comprise the activity pattern.

Time enters into the structure of activity patterns in two ways. The first is the duration of each activity, and the second is the time of occurrence of the activity. The duration of each activity is, of course, a basic ingredient in the account of a day's activities. At present little is known about the average amount of time spent on different activities, about how time spent varies among activities, or about how time spent on activities varies with characteristics of the persons performing them. The time of day when an activity occurs is also a basic ingredient in an account of a day's activities. But it also may be important to the duration, the choice, and the sequencing of activities.

It is possible that the duration of an activity influences whether that activity is chosen at all and how it is sequenced with other activities. This would apply to those activities that normally have a fixed minimum-time duration. The reciprocal influence suggested is that the sequence of activities chosen may influence the duration of each of the activities. The time of day may have an influence on the selection and linkage of activities. Some activities by custom or by their very nature are performed in the evening, for example, rather than in the morning.

The distribution of adapted space of the facilities for particular activities and the distribution of transportation and other facilities for interaction among adapted spaces obviously are interrelated with the choice of activities, and perhaps are related to the time dimensions of activities. As is well known, people are inhibited to some degree by distance, however measured. It follows that if facilities for a particular activity are relatively inaccessible, that activity probably will be performed infrequently. It may be that an activity will have a longer duration when access is difficult than when access is easy. If this is so, then the location of adapted spaces is relevant to both the occurrence and linkage of activities. The effective distribution of adapted spaces can be changed by modification of the transportation facilities for movement (the less tangible quality of them); this change may affect the selection and sequencing of activities. It should be pointed out that the distribution of activity places is viewed by the individual from the perspective of his location in the urban area, either in the conduct of some out-of-home activity or at home. Thus his view of the opportunities for various activities changes as he moves about the urban area. Because he often selects activities one at a time from home and after accomplishing them returns home, the residential location is the dominant focus from which the individual views the alternate opportunities for engaging in out-of-home activities. The distribution of adapted spaces and transportation facilities may affect either or both the time of day at which an activity is engaged in or the duration of the activity. Perhaps an activity that is difficult to perform because of some factor of space will, when it is chosen, have a longer duration than would otherwise be the case. If the preferred time for this activity is midday, it might be postponed until evening because sufficient time is not available at midday.

Activity choice itself can be explained only in terms of the motivation, needs, wants, and capabilities of the individual. The whole range of socioeconomic characteristics of the individual and the family unit of which he is a part is the source from which an explanation and structure for the variations and patterns in activity choice will be sought. The selection of activities by an individual may be a function of his preferences, tastes, information of alternatives, habits, or financial circumstances, and most certainly is a function of his requirements for personal and household maintenance. An understanding of activity patterns and linkages must be based on these factors, but it is an underlying hypothesis of the approach suggested here that the dimensions of time and space are also significant to the structuring of activity patterns.

In this discussion, these relationships between time, space, and activity are posited for the short run—a period of a day in a given urban environment in which the tastes, preferences, and attitudes are fixed for the moment. Looking to the long-run problems of projecting activity choices over time, it is clear that as preferences and attitudes change and as technology changes, the relationships will become more complex. A model that attempted to deal with activity choices over time would probably have to incorporate reciprocal relationships among the dimensions of time, space, and activity. As the urban area grows and changes, for example, prevalent choice patterns of activities will be reflected in the amount and location of adapted spaces designed to accommodate them.

PRELIMINARY EMPIRICAL RESULTS

As already indicated, two groups of activity linkages are being observed. First is the daily cycle; second is the composition of each journey made from the home during that cycle. Beginning with the widely used assumption that the majority of travel is home based—people start from home in the morning, return home, leave again, and

finally end up at home at night—a journey is defined here as a home-to-home circuit comprised of two or more trips (in the usual sense of the word "trips").

Data

The data used for this study are the "home interview" data obtained by the Niagara Frontier Transportation Study in the Buffalo, New York, area in 1962. The data represent the results of direct in-home interviews of 4 percent of 300,000 households in the study area. Along with considerable social, economic, demographic, and geographic data, each interview obtained a complete description of the out-of-home travel of the members of the household for a selected day. Information on activities has been derived from this trip information by inverting the original data. Activities have been defined in terms of combinations of the original trip purposes and land use at the origin and destination of each trip in the original data. For example, each land-use and trip-purpose combination, such as recreation in local parks or recreation at spectator sports, could be considered a separate activity. Many of the 400 possible activities are either too similar to others or occur too rarely to be treated separately. These have been combined into 43 distinct activity codes for this study (Table 1).

Additionally, three of the conventional trip-purpose codes—ride as a passenger, change mode (of travel), and serve passenger—are transportation-connected activities that are secondary to the real purpose or activity represented by the travel. To obtain purposeful trips in terms of activities, the transportation-connected purposes were eliminated by a process similar to linking of trips traditionally performed in transportation analysis. For example, two trips that were recorded as "home to change-mode" and "change-mode to work" respectively were combined in a single home-to-work trip.

The resulting data set contains information on the type, location, duration, and time of day of some 92,000 out-of-home activities performed by the 55,000 members of 16,000 households on a selected weekday. These have been reduced to a working data set of 10,300 households (for which trip reports are available) including 24,800 persons who made 33,500 journeys containing 83,300 trips. It should be emphasized that these data contain information only on out-of-home activities that require the use of transportation facilities—private car, taxi, bus, rail, or truck. The only walking trips included in the original data are those to work; no bicycle trips are recorded. These limitations are most severe when attempting to examine the activities of school children

TABLE 1

PERSON ACTIVITY CODES BASED ON TRIP PURPOSE AND LAND USE

Code	Activity	Code	Activity
01	Home	23	Social-recreation/indoor
02	Work/residential	24	Social-recreation/clubs
03	Work/retail	25	Social-recreation/schools, museums, libraries
04	Work/service and offices	26	Social-recreation/hospitals, etc.
05	Work/wholesale	27	Social-recreation/church
06	Work/durable manufacturing	28	Social-recreation/outdoor
07	Work/nondurable manufacturing	29	Social-recreation/local parks
08	Work/institutional	30	Social-recreation/spectator sports
09	Work/recreation	31	Social-recreation/miscellaneous
10	Work/transportation terminals and facilities	32	Eat meal/residential land
11	Work/other	33	Eat meal/restaurant or club
12	Shop/food, drug, liquor	34	Personal business/residential land
13	Shop/other convenience goods	35	Personal business/personal services
14	Shop/department store	36	Personal business/medical, dental
15	Shop/other shopping goods	37	Personal business/business service
16	Shop/automotive	38	Personal business/other services
17	Shop/miscellaneous other	39	Personal business/manufacturing and wholesaling
18	School/elementary	40	Personal business/hospitals, etc.
19	School/secondary	41	Personal business/church
20	School/other, including college	42	Personal business/other public buildings
21	Social-recreation/residential land	43	All other personal business
22	Social-recreation/eating and drinking		

and teenagers. Also missing are the walking mode shopping trips in the older, dense neighborhoods near the city core, and walking trips from one store to another in a shopping center and the heart of the central business district (CBD).

Measures Describing Activities

Several measures have been evolved in examining activity patterns and linkages. Because the primary purpose in analyzing and attempting to simulate activity patterns is to determine how people actually use the urban area and how they will use it in the future, we first are trying to discover who links activities, to what extent, and under what conditions. To date, the "who does it" and "to what extent" questions have been examined. The question of under what conditions it is done is much more complex because it involves not only the individual's schedule of activity demands, but the supply (the amount and location) of opportunities for satisfying them.

To analyze activity linkages within journeys and over a 24-hour day, the trips to and from home and activities and between activities are structured as an absorbing Markov chain in which the return-home state is absorbing and all the other states—leaving home and all the out-of-home activities—are nonabsorbing. The logic of treating travel/activity data this way is, as argued before, that each out-of-home journey is a "closed loop" containing one or more activities and that the whole day's travel is simply a series of such loops. The benefit of using the absorbing-Markov-chain model for analysis of the data is simply that it permits easy and economical reconstruction of journey patterns from the source data; and by taking advantage of some of the properties of Markov processes, we can gain some additional information from our data. [For further discussion see Kemeny et al. (4).]

A fundamental matrix is derived from the summed absorbing chains (representing the probability of visiting activity states during a day). The first row of the matrix represents the average number of activities visited given that the leave-home state has been exited. Each entry in the first row, in turn, represents the number of times the activity will be visited. The effect is to provide a summary of the entire original matrix for the leave-home condition.

Other measures are more traditional: averages of time spent in various activities, numbers of journeys per day by households, and frequency of occurrence of various activities by time of day.

Journey Complexity

Figure 1 shows the average number of trips per journey for households classified by four variables. The mean, 2.486, is indicated by the single horizontal line, and the dotted lines are two standard deviations from the mean. It is apparent the high-income suburban whites make more complex journeys than do any other groups. Family size has a notable inverse relationship to journey complexity.

From other analyses it appears that members of households who live in apartments or two-family houses are less likely to make multiple activity journeys than are persons

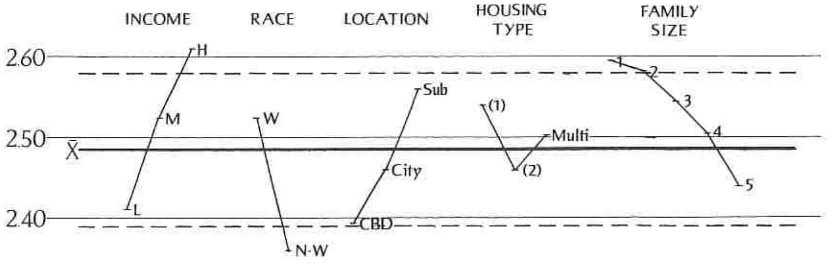


Figure 1.

living in single-family houses. The likelihood of multiple activity journeys by all members of a household is related inversely to the age of the head of the household; the older the head, the fewer non-home-based trips there are. Also, the likelihood of multiple activity journeys is related directly to household car ownership. If a car is owned by the household a person is more likely to link activities than if no car is owned; if two cars, a person is more likely to link activities than if only one car is owned. Similarly the likelihood of multiple activity journeys is related directly to occupation of the head of the household, ranked from high socioeconomic status to lower socioeconomic status. In summary, members of households that are young, white, middle class, live in single-family houses, and own several cars are most likely to link activities in complex out-of-home journeys. Members of households whose characteristics are the opposite of these are least likely to make multiple journeys. More of their out-of-home activities are done one at a time with a return home before another is undertaken.

From one perspective, multiple activity journeys would seem more economical and efficient than single activity journeys. It would seem reasonable to expect the poor and car-less to economize on trips by linking out-of-home activities. From another perspective, however, activity linkage itself can be viewed as a fairly luxurious practice involving comparison shopping, a desire for full utilization of available opportunities, and the ability to exercise these interests. We are far from reaching satisfactory explanations of this behavior, but we will not have the full answer until we can examine this behavior in terms of the location pattern of activity opportunities available.

Table 2 gives the estimated number of trips per journey for the same variables used in Figure 1. Several relationships among these control variables and the number of trips in a journey are evident. Considering the white middle-income group only, the relation of family size to journey length can be seen most clearly. As family size increases, journey length gets smaller; i.e., fewer activities are linked on a single journey or there is less likelihood of multiple activity journeys. For low- and high-income whites, family size seems to have little effect on journey length. For nonwhites there is no clear relation between family size and journey length.

The effect of race on journey length is also clearest for the middle-income group. In general, nonwhites make shorter journeys, and are less likely to link activities than are whites. The effect of income on journey length is clear for all groups in the table. The pattern is that the rich tend to make longer journeys than the poor. This pattern is subdued somewhat by the effect of family size and location. Residential location also is related to journey length. Suburban dwellers make longer journeys and are more likely to link activities than are central-city dwellers. This effect is least

TABLE 2

ESTIMATED ACTIVITIES PER JOURNEY BY HOUSEHOLD CHARACTERISTICS[a]

Income	Location	Activities per Journey by Family Size and Race									
		1		2		3		4		5+	
		White	Non-white	White	Non-white	White	Non-white	White	Non-white	White	Non-white
High (≥$10,000)	City core	2.600	—	2.869	2.083	2.557	3.000	2.344	2.406	2.477	2.25
	Rest of central city	2.100	—	2.856	—	2.458	—	2.538	2.667	2.405	—
	Suburbs	2.611	—	2.874	2.000	2.651	—	2.566	2.000	2.583	2.000
Middle ($5,000-9,999)	City core	2.727	2.333	2.508	2.443	2.429	2.542	2.367	2.264	2.364	2.295
	Rest of central city	2.611	2.000	2.510	2.000	2.486	2.433	2.465	2.000	2.409	2.000
	Suburbs	2.858	—	2.625	2.000	2.631	2.833	2.572	—	2.448	2.500
Low (<$5,000)	City core	2.296	2.429	2.388	2.697	2.355	2.234	2.233	2.260	2.242	2.188
	Rest of central city	2.478	—	2.458	2.286	2.362	2.193	2.400	2.000	2.344	2.100
	Suburbs	2.719	—	2.497	2.286	2.457	2.400	2.401	2.301	2.293	2.700

[a]From fundamental matrix based on observed behavior.

marked for high-income households because the effect of incomes is to increase journey length. These results support and reinforce those reported above. In summary, the members of rich, white, suburban households tend to link activities on complex journeys. The poor, nonwhite, city dwellers do not.

Activity Duration

Looking at the duration of activities during the day (Table 3), there is a clear difference in mean duration of activities when separated into those performed in the morning, afternoon, or evening. As was hypothesized in structuring the framework for analysis and simulation of activities, duration of activity is related to the time of day.

Activities last longest if they are performed (or at least started) in the morning. Excluding work, 20 of the remaining 31 activities given in Table 3 have their longest durations in the morning. The exceptions are primarily social-recreation and personal-business activities. Slightly more of the activities (17 versus 14) have longer duration in the afternoon than in the evening.

Skimming over the results in Table 3, some well-known behavior patterns are clearly shown as well as some of the limitations of the data. For example, the longest

TABLE 3

MEAN DURATION OF ACTIVITIES WITH DIFFERENT STARTING TIMES[a]

Code	Activity	Mean Duration (hr)			
		Independent of Start Time	Before 12 Noon	Between 12 Noon and 5 p.m.	After 5 p.m.
02	Work/residential	2.96	4.48	1.04	3.74
03	Work/retail	5.44	7.24	3.48	3.13
04	Work/service and offices	5.46	6.84	2.90	2.27
05	Work/wholesale	5.62	6.84	2.50	0.95
06	Work/durable manufacturing	7.68	8.22	5.88	9.47
07	Work/nondurable manufacturing	7.38	8.19	4.73	6.01
08	Work/institutional	6.05	7.10	3.35	2.90
09	Work/recreation	5.45	8.93	4.03	2.75
10	Work/transportation terminals and facilities	6.59	8.01	3.47	5.10
11	Work/other	6.57	7.34	5.20	—
12	Shop/food, drug, liquor	0.54	0.78	0.53	0.47
13	Shop/other convenience goods	1.35	3.56	0.28	0.91
14	Shop/department store	1.16	1.62	1.16	0.97
15	Shop/other shopping goods	0.53	0.60	0.50	0.53
16	Shop/automotive	1.21	1.04	0.82	1.68
17	Shop/miscellaneous other	1.16	0.80	0.64	3.07
18	School/elementary	6.19	6.47	4.13	—
19	School/secondary	7.35	7.60	2.50	1.10
20	School/other, including college	4.58	5.07	—	2.23
21	Social-recreation/residential land	2.74	4.95	3.32	2.04
22	Social-recreation/eating and drinking	1.18	2.01	1.51	1.00
23	Social-recreation/indoor	3.00	—	3.82	2.82
24	Social-recreation/clubs	2.35	2.86	1.94	2.35
25	Social-recreation/schools, museums, libraries	2.31	4.60	1.03	2.50
26	Social-recreation/hospitals, etc.	1.10	—	1.23	0.98
27	Social-recreation/church	2.84	1.68	4.68	2.78
28	Social-recreation/outdoor	2.83	4.96	2.54	1.71
29	Social-recreation/local parks	2.68	6.80	3.04	2.00
30	Social-recreation/spectator sports	3.74	9.88	3.81	2.85
31	Social-recreation/miscellaneous	2.25	1.40	1.89	2.50
32	Eat meal/residential land	1.70	0.45	1.80	1.49
33	Eat meal/restaurant or club	1.05	1.01	1.13	0.99
34	Personal business/residential land	1.56	0.90	1.19	1.91
35	Personal business/personal services	1.04	1.11	0.95	1.18
36	Personal business/medical, dental	1.33	1.66	1.55	0.86
37	Personal business/business service	0.61	0.66	0.55	0.69
38	Personal business/other services	1.29	2.40	0.83	0.91
39	Personal business/manufacturing and wholesaling	0.81	1.18	0.58	0.47
40	Personal business/hospitals, etc.	2.30	9.03	2.92	0.93
41	Personal business/church	1.94	2.81	1.07	1.47
42	Personal business/other public buildings	1.02	1.48	0.56	0.92

[a]Based on 10 percent sample of full data set.

shopping activity is the morning shopping excursion for miscellaneous convenience goods—the housewife's tour. The big shopping excursion for a new car, on the other hand, occurs in the evening. The long evening shopping tour for miscellany probably reflects the activities of those who were "out shopping around" at shopping centers or downtown. Perhaps these might be better classified as recreation activities, at least in part.

Activity Choice and Time of Day

In the conceptual framework it is hypothesized that choice of activity is related to time of day. As given in Table 4, frequency of occurrence of activities undertaken in the morning, afternoon, and evening are about the same. But many more social-recreation activities occur in the evening than in the morning or afternoon; more shopping is done in the afternoon and evening than in the morning (although morning shopping activities typically have longer duration); and more personal business activities occur in the afternoon than in the evening or morning. To validate the hypothesis that activity choice is related to time of day, the following were used: (a) the proposition

TABLE 4

FREQUENCY OF OCCURRENCE OF ACTIVITIES[a]

Code	Activity	Number Occurring			
		Over 24-Hour Period	Start Before 12 Noon	Start Between Noon and 5 p.m.	Start After 5 p.m.
02	Work/residential	137	60	56	21
03	Work/retail	257	136	102	19
04	Work/service and offices	353	233	97	23
05	Work/wholesale	88	64	22	2
06	Work/durable manufacturing	324	231	81	12
07	Work/nondurable manufacturing	194	145	39	10
08	Work/institutional	127	92	30	5
09	Work/recreation	12	4	6	2
10	Work/transportation terminals and facilities	127	83	32	12
11	Work/other	13	10	2	1
12	Shop/food, drug, liquor	629	83	285	261
13	Shop/other convenience goods	39	11	19	9
14	Shop/department store	279	44	134	101
15	Shop/other shopping goods	131	22	51	58
16	Shop/automotive	33	7	13	13
17	Shop/miscellaneous other	36	11	18	7
18	School/elementary	137	121	16	—
19	School/secondary	135	129	4	2
20	School/other, including college	25	21	1	3
21	Social-recreation/residential land	578	58	184	336
22	Social-recreation/eating and drinking	119	9	23	87
23	Social-recreation/indoor	85	1	17	67
24	Social-recreation/clubs	31	7	9	15
25	Social-recreation/schools, museums, libraries	47	4	12	31
26	Social-recreation/hospitals, etc.	21	—	10	11
27	Social-recreation/church	52	4	4	44
28	Social-recreation/outdoor	49	11	23	15
29	Social-recreation/local parks	77	7	18	52
30	Social-recreation/spectator sports	49	5	9	35
31	Social-recreation/miscellaneous	70	7	16	47
32	Eat meal/residential land	55	2	44	9
33	Eat meal/restaurant or club	178	25	77	76
34	Personal business/residential land	125	12	43	70
35	Personal business/personal services	60	13	33	14
36	Personal business/medical, dental	75	15	34	26
37	Personal business/business service	123	40	62	21
38	Personal business/other services	94	26	40	28
39	Personal business/manufacturing and wholesaling	24	10	10	4
40	Personal business/hospitals, etc.	41	3	16	22
41	Personal business/church	36	15	8	13
42	Personal business/other public buildings	94	41	37	16

[a]Based on 10 percent sample of full data set.

that the probabilities of home-activity, activity-activity, and activity-home trips were equal for different times of the day tested; (b) chi-square; and (c) analysis of variance tests. In virtually every case the test showed that the probabilities of transition occurrence were significantly different by time of day.

Activity Profile

Figures 2 through 6 show the profile of average numbers of visits to various activities (from the fundamental of the transition matrix), given that the initial leave-home state has been exited. (The absorbing return-home state is omitted; by definition it is always equal to 1.0 per journey.) Another way of looking at these is to read the figures as the number of visits per 1,000 journeys. For the total white population in the sample (Fig. 2, lower) the 1,000 journeys will include 2,513 trips, 1,523 to activities and 1,000 back home. Of the 1,523 trips, about 180 will be made to shop at a food, drug, or liquor store.

The major noticeable difference in these profiles is that white, higher income families have a more even distribution of activities than do their opposite numbers. Apparently, as observed with journey complexity—or perhaps interacting with journey complexity—higher income, white suburbanites have more diverse activity choice. Whites exhibit no strong orientation in work selection, whereas nonwhites are concentrated to some extent. Nonwhites make most of their social-recreational trips (in this activity breakdown) to other residential land. Although both show high frequencies for shopping at food, drug, and liquor stores, whites make many more such trips. Whites also go out to eat at restaurants and clubs more often than nonwhites.

Similar trends are evident in the three income groups and the three residential locations. Unanswered is the obvious question of intercorrelation of race, income, and residential location; further analysis is needed to separate these effects. [Extension of multivariate statistical analyses to the vector representation of activity choice is conceptually not too difficult, but is computationally tedious. See Anderson (5) for further discussion.] As would be anticipated, housing type shows similar differences (Fig. 5). In regard to family size, the concentration in a few activities is most intense for single persons and decreases for larger households (Fig. 6).

The implications of the evident concentration of a few specific activities (and the nature of the concentration) are similar to those that might be anticipated from the fundamental-supplemental division suggested by Chapin. Low-income nonwhite families

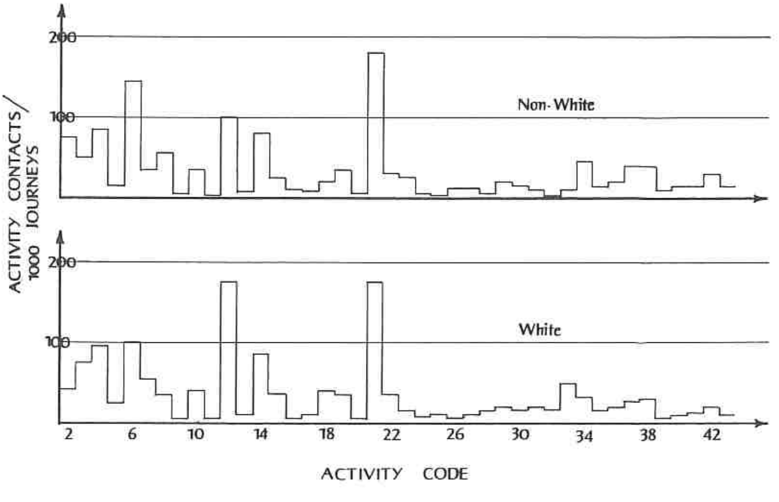


Figure 2.

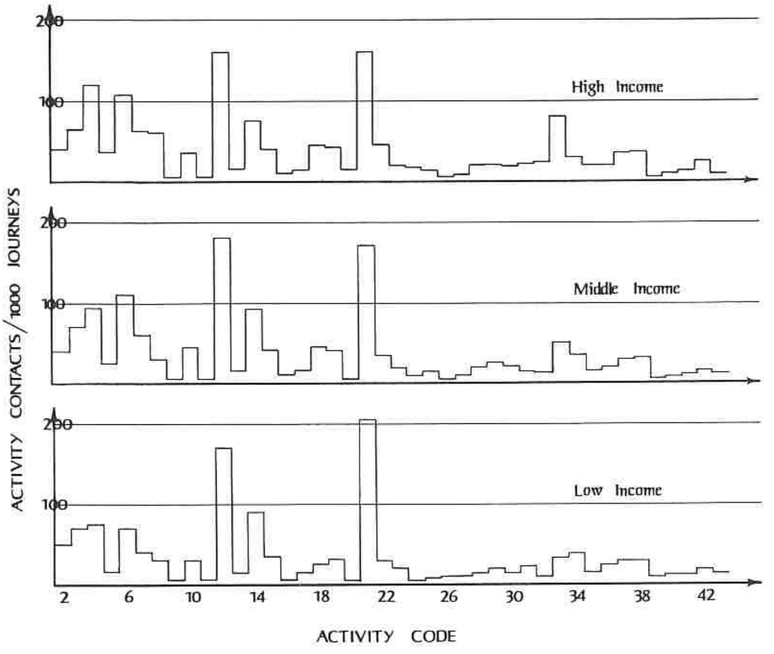


Figure 3.

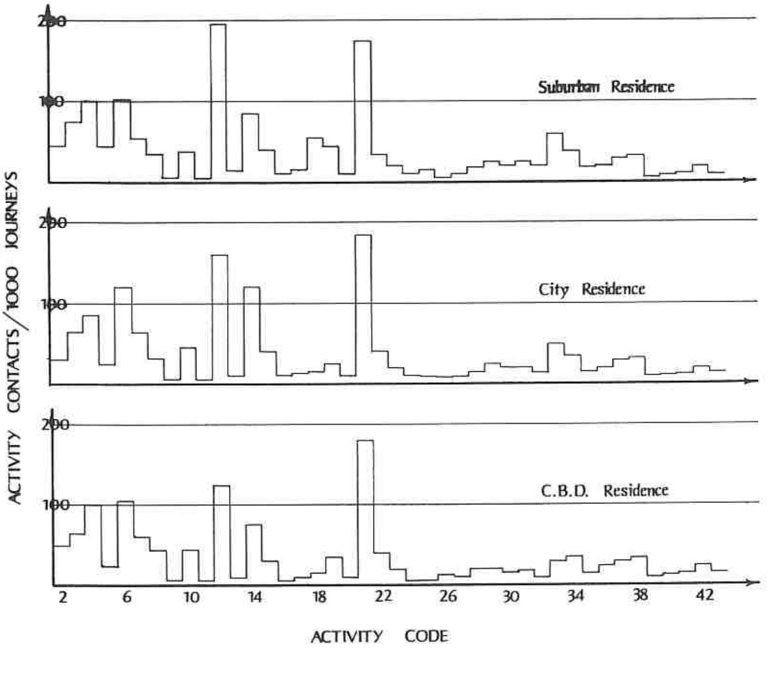


Figure 4.

There are strong repetitive patterns evident in the analysis reported here, with some anticipated variation, not all of which is easily explained. It is hoped that this material, drawn from the rich data base of home interview surveys, may be linked with more intensive time budget data to fill the gaps in home interview travel information. The eventual product should be a stronger insight into the demand for the facilities available in the urban structure: not only transportation facilities, but also those meeting the basic needs of families in their normal routine and satisfying their demands for leisure.

ACKNOWLEDGMENT

The research reported here is supported by a grant from the National Science Foundation. That support is gratefully acknowledged.

REFERENCES

1. Chapin, F. Stuart, Jr. Activity Systems and Urban Structure: A Working Schema. Jour. Amer. Inst. Planners, Vol. 34, No. 1, 1968, pp. 11-18.
2. Chapin, F. Stuart, Jr., and Hightower, H. C. Household Activity Systems—A Pilot Investigation. Center for Urban and Regional Studies, Univ. of North Carolina, Chapel Hill, 1966.
3. Chapin, F. Stuart, Jr. Urban Land Use Planning. 2nd Ed., Univ. of Illinois Press, Urbana, 1965.
4. Kemeny, John B., et al. Finite Mathematical Structures. Prentice-Hall, Englewood Cliffs, 1959.
5. Anderson, T. W. An Introduction to Multivariate Statistical Analysis. John Wiley and Sons, New York, 1958.
6. Hamburg, J. R., Guinn, C. R., Lathrop, G. T., Hemmens, G. C. Linear Programming Test of Journey-to-Work Minimization. Highway Research Record 102, 1965, pp. 67-75.

THE NATIONAL ACADEMY OF SCIENCES is a private, honorary organization of more than 700 scientists and engineers elected on the basis of outstanding contributions to knowledge. Established by a Congressional Act of Incorporation signed by Abraham Lincoln on March 3, 1863, and supported by private and public funds, the Academy works to further science and its use for the general welfare by bringing together the most qualified individuals to deal with scientific and technological problems of broad significance.

Under the terms of its Congressional charter, the Academy is also called upon to act as an official—yet independent—adviser to the Federal Government in any matter of science and technology. This provision accounts for the close ties that have always existed between the Academy and the Government, although the Academy is not a governmental agency and its activities are not limited to those on behalf of the Government.

The NATIONAL ACADEMY OF ENGINEERING was established on December 5, 1964. On that date the Council of the National Academy of Sciences, under the authority of its Act of Incorporation, adopted Articles of Organization bringing the National Academy of Engineering into being, independent and autonomous in its organization and the election of its members, and closely coordinated with the National Academy of Sciences in its advisory activities. The two Academies join in the furtherance of science and engineering and share the responsibility of advising the Federal Government, upon request, on any subject of science or technology.

The NATIONAL RESEARCH COUNCIL was organized as an agency of the National Academy of Sciences in 1916, at the request of President Wilson, to enable the broad community of U.S. scientists and engineers to associate their efforts with the limited membership of the Academy in service to science and the nation. Its members, who receive their appointments from the President of the National Academy of Sciences, are drawn from academic, industrial, and government organizations throughout the country. The National Research Council serves both Academies in the discharge of their responsibilities.

Supported by private and public contributions, grants, and contracts, and voluntary contributions of time and effort by several thousand of the nation's leading scientists and engineers, the Academies and their Research Council thus work to serve the national interest, to foster the sound development of science and engineering, and to promote their effective application for the benefit of society.

The DIVISION OF ENGINEERING is one of the eight major Divisions into which the National Research Council is organized for the conduct of its work. Its membership includes representatives of the nation's leading technical societies as well as a number of members-at-large. Its Chairman is appointed by the Council of the Academy of Sciences upon nomination by the Council of the Academy of Engineering.

The HIGHWAY RESEARCH BOARD, an agency of the Division of Engineering, was established November 11, 1920, as a cooperative organization of the highway technologists of America operating under the auspices of the National Research Council and with the support of the several highway departments, the Bureau of Public Roads, and many other organizations interested in the development of transportation. The purpose of the Board is to advance knowledge concerning the nature and performance of transportation systems, through the stimulation of research and dissemination of information derived therefrom.

HIGHWAY RESEARCH BOARD
NATIONAL ACADEMY OF SCIENCES—NATIONAL RESEARCH COUNCIL
2101 Constitution Avenue Washington, D. C. 20418

ADDRESS CORRECTION REQUESTED